AD-A092 708 AIR FORCE GEOPHYSICS LAB HANSCOM AFB MA F/G 3/2

STATISTICAL METHODS FOR SOLAR FLARE PROBABILITY FORECASTING.(U)

SEP 80 D F VECCHIA, P V TRYON, G A CALDWELL

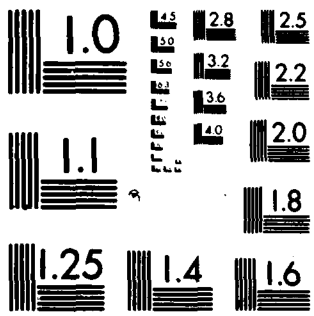
UNCLASSIFIED AFGL-TR-80-0336 NL

1 of 1

END

DATE FILMED

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963 A

AFGL-TR-80-0336

STATISTICAL METHODS FOR SOLAR FLARE PROBABILITY FORECASTING

Dominic F. Vecchia
Peter V. Tryon
Ginger A. Caldwell
Richard H. Jones

Statistical Engineering Division
Center for Applied Mathematics
National Engineering Laboratory
National Bureau of Standards
Boulder, Colorado 80303

Final Report
June 1979 - February 1980

September 1980

Approved for public release; distribution unlimited

Prepared for NOAA Space Environment Laboratory, U.S. Air Force

AIR FORCE GEOPHYSICS LABORATORY
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
HANSCOM AFB, MASSACHUSETTS 01731

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

14

1. ~~REPORT NUMBER~~ AFGL-TR-80-0336

2. GOVT ACCESSION NO. AD-A092 708

3. RECIPIENT'S CATALOG NUMBER

6

4. TITLE (and Subtitle) STATISTICAL METHODS FOR SOLAR FLARE PROBABILITY FORECASTING.

5. TYPE OF ~~REPORT & PERIOD COVERED~~ Scientific Final rept. Jun[redacted]79 - Feb[redacted]80

6. ~~PERFORMING ORG. REPORT NUMBER~~

10

7. ~~AUTHOR(s)~~ Dominic F. Vecchia, Richard H. Jones, Peter V. Tryon, Ginger A. Caldwell

8. CONTRACT OR GRANT NUMBER(s) Project Order ESD 9-0959 to National Oceanic and Atmospheric Administration

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Statistical Engineering Division
Center for Applied Mathematics
National Engineering Laboratory
National Bureau of Standards
Boulder, Colorado 80303 (U.S.A.)

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
61102F
2311G310 17 G3

11. CONTROLLING OFFICE NAME AND ADDRESS
Air Force Geophysics Laboratory
Space Physics Division
Hanscom AFB, MA 01731
Monitor/D.F. Neidig/PH

11

12. ~~REPORT DATE~~ Sep[redacted]80

13. NUMBER OF PAGES 80 12

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report) Unclassified

15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

*Richard J. Jones, Department of Bimetrics, University of Colorado Medical School, Denver CO., is a mathematical statistician affiliated with the Statistical Engineering Division, NBS.

Prepared for NOAA Space Environment Laboratory, U.S. Air Force

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

solar flare; probability forecasting; discriminant analysis; logistic regression; multiple linear regression; Brier score.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The Space Environment Services Center (SESC) of the National Oceanic and Atmospheric Administration provides probability forecasts of regional solar flare disturbances. This report describes a statistical method useful to obtain 24 hour solar flare forecasts which, historically, have been subjectively formulated.

In Section 1 of this report flare classifications of the SESC and the particular probability forecasts to be considered are defined. In Section 2 we

DD FORM 1 JAN 73 1473

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

4

Item 20 continued.

describe the solar flare data base and outline general principles for effective data management. Three statistical techniques for solar flare probability forecasting are discussed in Section 3, viz, discriminant analysis, logistic regression, and multiple linearregression. We also review two scoring measures and suggest the logistic regression approach for obtaining 24 hour forecasts. In Section 4 a heuristic procedure is used to select nine basic predictors from the many available explanatory variables. Using these nine variables logistic regression is demonstrated by example in Section 5. We conclude in Section 6 with broad suggestions regarding continued development of objective methods for solar flare probability forecasting.

Table of Contents

		Page
1.	INTRODUCTION	1
2.	REGION ANALYSIS DATA	3
2.1	Description of Data	3
2.2	Recommendations for Data Management	6
3.	SOLAR FLARE FORECASTING METHODS	6
3.1	Probability Forecast Methods	7
3.1.1	Discriminant Analysis	7
3.1.2	Logistic Regression	9
3.1.3	Extension to Polychotomous Case	9
3.1.4	Multiple Linear Regression	10
3.2	Scoring Probability Forecasts	11
3.3	Discussion	12
3.4	Recommendations	13
3.4.1	Probability Forecasting	13
3.4.2	Estimation of Flare Intensity	14
4.	SELECTION OF PREDICTORS	15
4.1	Transformation of Variables	15
4.2	Selection of Basic Variables	16
4.3	Variable Rescaling	21
5.	EXAMPLE	22
5.1	Case Selection	23
5.2	Grouping of Cases	23
5.3	Validation	24
5.4	Parameter Estimation	24
5.5	Model Evaluation	29
6.	SUMMARY AND RECOMMENDATIONS	34
7.	ACKNOWLEDGEMENT	35
	REFERENCES	36
APPENDIX A:	Region Analysis Variables	38
APPENDIX B:	Probability Forecast Tables	44
APPENDIX C:	Crosstabulations	58

Tables

		Page
1.	Flare Classification by X-ray Yield	1
2.	Region Analysis Variables	4
3.	Number of Non-Missing Observations	17
4.	Linear Trend "F" Statistics	18
5.	Means and Standard Deviations--Set 1	19
6.	Means and Standard Deviations--Set 2	20
7.	Means and Standard Deviations--Set 3	21
8.	Forecast Models	23
9.	Case Grouping	23
10.	Overall Chi-Square Significance Levels	25
11.	Estimated Parameters--Model I	26
12.	Estimated Parameters--Model II	27
13.	Estimated Parameters--Model III	28
14.	Estimated Parameters--Model IV	29
15.	Brier Scores for $Pr[C, M \text{ or } X \text{ Flare} \mid \underline{x}]$	31
16.	Brier Scores for $Pr[M \text{ or } X \text{ Flare} \mid \underline{x}]$	32
17.	Brier Scores for $Pr[X \text{ Flare} \mid \underline{x}]$	33

1. INTRODUCTION

Historically, solar flare forecasting methods have been subjectively formulated, relying heavily on forecaster insight. This report addresses the desire for an objective technique for solar flare probability forecasting, in light of the importance of accurate forecasts to the scientific community and the general public.

The Space Environment Services Center (SESC), a part of the NOAA Space Environment Laboratory in Boulder, Colorado, provides 24-hour probability forecasts of regional solar flare disturbances. Variables comprising predictive information for this subjective method are those found or conjectured to be useful by SESC forecasters (Hirman and Flowers (1979)). The "region analysis" variables thought essential to flare occurrence serve, as well, for our development of an objective technique. For a complete discussion of goals and services of the SESC, the reader is referred to Heckman (1979b) and Mangis (1975).

Solar flare forecasts made by the SESC predict both the occurrence and magnitude of flares. Four classes, denoting the largest event in a 24-hour period, can be identified: (1) no flare, (2) class C flare, (3) class M flare and (4) class X flare. Ranges of X-ray yield defining flare classes are listed in Table 1 (see Mangis (1975)).

Table 1. Flare Classification by X-Ray Yield

Class	Energy Output E in the 1-8 Å Spectral Range
C	$10^{-6} \leq E < 10^{-5}$ W/m^2
M	$10^{-5} \leq E < 10^{-4}$ W/m^2
X	$10^{-4} < E$ W/m^2

An additional coefficient appended to the letter designator indicates the relative intensity within the appropriate energy range. For example, an X-ray class M3 flare would yield 3×10^{-5} W/m^2, an X5 would yield 5×10^{-4}, and so on. A non-energetic flare is one which is less than class C1, that is, one which produces less than 1×10^{-6} W/m^2. It is the moderate class M flare and the major class X flare which are of greatest consequence to the near-earth environment.

If we let Z represent a random variable such that Z = 0, 1, 2, or 3 corresponding to the largest flare which occurs in the next 24 hours in a selected region, then estimates are provided by the SESC for the following conditional probabilities:

$$
\begin{array}{ll}
\text{(i)} & \Pr[Z=0\,|\,\underline{x}] \\
\text{(ii)} & \Pr[Z\geq 1\,|\,\underline{x}] \\
\text{(iii)} & \Pr[Z\geq 2\,|\,\underline{x}] \\
\text{(iv)} & \Pr[Z=3\,|\,\underline{x}]
\end{array} ,
$$

where $\underline{x}$ denotes an observed vector of prediction variables associated with the selected region.[1]

Objective prediction of probabilities (i)-(iv), or variants of these, has been accomplished with some success by Hirman, et al. (1980) using the technique of multivariate discriminant analysis and by Vecchia, et al. (1980) using logistic regression. These results demonstrate potential improvement on subjective forecasts and indicate that, perhaps, the time has arrived for an expanded effort to develop an objective technique. This would enable forecasters to attach quantitative significance to the many interrelated variables comprising the inputs to any forecasting method.

In this report we propose the technique of logistic regression for prediction of (i)-(iv). Though we undertake to examine only 24-hour forecasts, with continued development and understanding the procedure can be applied to other time frames. It should be emphasized that a complete evaluation of any proposed technique can result only from comparison to the baseline measure provided by the subjective forecasting system. For the example we provide scores for intercomparison of logistic regression, discriminant analysis, and the SESC forecasts, based on measures already used by the SESC.

In section 2 we describe the current solar flare data base and outline a general data management procedure which is essential if present and future records are to provide the statistical information of interest and importance. Section 3 is a brief review of three techniques which have been suggested to forecast solar flares --discriminant analysis, logistic regression, and multiple linear regression. We also review two scoring measures and discuss our preference for the logistic regression approach. A heuristic procedure for selection and transformation of variables is employed in section 4 to obtain nine basic variables for the examples is section 5. We conclude in section 6 with broad suggestions regarding continued development of objective methods for solar flare probability forecasting.

---

[1]The notation "$Z=0|\underline{x}$" is read "Z=0 given $\underline{x}$." For example, (iii) is the probability that an M or X flare will occur in the next 24 hours given the predictors $\underline{x}$.

2. REGION ANALYSIS DATA

2.1 Description of Data

The data result from a data collection and analysis scheme initiated by the SESC on January 1, 1977. These observations, collected from SESC sensors and from cooperating agencies and institutions, reflect the complexity, magnetic configuration, age, location, and past history of active solar regions. All explanatory variables included in the SESC record are available in near real time though some are not accessible on a daily basis. Any use of the SESC data base to develop solar flare forecasting techniques should acknowledge this limitation.

Data collected by the SESC are divided into four categories: white light, H-alpha, radio, and region history. In addition, the variables are mixed--continuous and discrete--and some are dichotomous. A complete list of 48 variables and brief descriptions of each is provided in Appendix A. Some variables are recoded and/or reordered versions of the original SESC observations. For the most part, this recoding was based on a reassessment by staff forecasters of the relation of the variables to solar flare activity.

An abridged list of available information is presented in Table 2. Variables listed are those considered in the current study and do not include information on the location of active regions on the sun. It is indicated if a variable is continuous or discrete and, if discrete, the number of distinct levels assumed. Also noted is the source of each variable. Additional information in Table 2 reflects the fact that variables range from completely objective measurements to highly subjective forecaster observations, such as visual evaluation of optical telescope photographs. This range of objectivity has been coded into three categories by SESC forecasters.

The current study utilizes 6097 region-day records collected from January 1, 1977 to January 31, 1979. These record were reduced to 4487 records by the elimination of records indicating the absence of sunspots, since such regions rarely produce flares. For these cases, two-way crosstabulations of FLARER (Variable 39) with other variables are given in Appendix C.

Table 2. Region Analysis Variables

Number[1]	Name	Type[2]	Source[3]	Description[4]
1	DATE	D	SESC	Year, month, day (3)
8	AGE	D-15	SESC	Age of region (3)
10	MAGCLAS	D-7	SESC	Magnetic class (2)
11	RV	D-3	MW	Magnetic field strength polarity (3)
12	MAGSTR	D-98	MW	Magnetic field strength (3)
13	MAGGRAD	C	MW	Magnetic gradient in gamma/km (3)
14	SSDYNAM	D-4	SOON	Sunspot dynamics (1)
15	SSINTER	D-2	SESC	Interaction with another region (1)
16	STGDEV	D-6	SESC	Stage of development (2)
19	SECTEOW	D-8	SESC	Relationship with nearest sector boundary (3)
20	PLAGFIL	D-6	SOON	Plage compactness and embedded filament (1)
21	NEUTLOR	D-5	SOON	Main neutral line orientation within plage (1)
22	REVPOL	D-2	MW	Orientation within plage (3)
23	NEUTLCOM	D-5	SOON	Neutral line complexity (1)
24	NEUTLCHG	D-3	SOON	Neutral line temporal changes (1)
25	ASSOCFIL	D-5	SOON	Associated filament (2)
26	BRTPTS	D-3	SOON	Bright points (3)
27	PLAGFLUX	D-2	SOON	Plage fluctuations (3)
28	ISOPOLE	D-2	SOON	Isolated Pole (2)
29	EFR	D-3	SOON	Emerging flux (2)
30	AFS	D-2	SOON	AFS (3)
33	FIRSTAPP	D-7	SESC	Regions first appearance (3)
35	CRFOR	C	SESC	C flare forecast for region
36	MRFOR	C	SESC	M flare forecast for region
37	XRFOR	C	SESC	X flare forecast for region
39	FLARER	D-4	SESC	Largest flare in next 24 hours
41	FLUX	C	SESC	10 cm flux (3)
46	FLARERT	D-4	SESC	Largest flare today (3)
47	RECSPOT	D-8	SESC	Recoded sunspot class (2)

[1]Variable numbers correspond to those of the complete data list in Appendix A.
[2]Variable type codes are: C = continuous; D-n = discrete-number of levels.
[3]Sources are: Space Environment Services Center (SESC); Solar observing optical network (SOON); Mt. Wilson (MW), Boulder, CO.
[4]Description parenthetic codes denote level of objectivity for the measurement: 1 = least objective, 2 = moderately objectively, 3 = most objective.

2.2 Recommendations for Data Management

From the outset of this study it was apparent that data access and retrieval difficulties would arise, though the actual extent of problems was not anticipated. Through September 30, 1979, four blocks of data representing 8001 SESC region-day cases from January 1, 1977, through June 30, 1979, have been transferred for analysis to the NOAA CDC 6600 computer. Following substantial delays for coding and keypunching blocks 1-3 (Jan. 77-Dec. 77; Jan. 78-June 78; July 78-Jan. 79), it was recommended that the SESC develop capability for the direct transfer of data from local terminals in near real time --perhaps on a monthly basis. The necessary software was tested in the transfer of the fourth data block, consisting of 1904 data cases from February 1, 1979 through June 10, 1979.

Preliminary examination of the data revealed many inconsistencies and/or errors. To the extent possible, errors detected in blocks 1-3 were corrected by SESC staff members. In some cases, impossible values were recoded as missing, resulting in a loss of information. Extensive and serious problems with data set 4, unless they are resolved, will prohibit any analysis which could be expected to provide reliable information.

<u>The importance of data collection and management cannot be exaggerated.</u> The difficulties encountered in the course of the present study will preclude a rigorous and completely reliable analysis of information contained in the data base. Such basic problems, if not satisfactorily resolved, may result in future records which cannot possibly provide the statistical information of interest or importance. To accomplish a reliable, near real-time data transfer and analysis scheme will require the following important components:

1. A reliable on-line procedure for coding, recording, and local storage of region analysis data cases.

2. An error free software system for the (direct) periodic transfer of region analysis data to the larger computer system required for complex statistical analyses. This package should provide a complete data record, including variables created or recoded from original variables.

3. Local (SESC) magnetic tape storage of original data sets until the complete verification of a successful transfer to the larger computer is accomplished.

4. Magnetic tape storage of data sets on the larger computer at the time successful transfer is verified. This should be in duplicate if possible, because of the importance and size of the data base.

5. An efficient software package for data base management on the large computer. The software should allow listing and editing of records. Programs to scan data for detectable inconsistencies should be developed and used regularly. To date, the SPSS statistical programs package (Nie, et al. (1975)) has been employed for data management, and has the additional advantage of providing descriptive and analytic statistical procedures useful for studying the solar flare data.

It should be emphasized that attention unnecessarily devoted to matters of data management and, in particular, to correction of recording or transfer errors, will postpone reliable statistical analyses. In all, it would be fair to estimate that a major portion of the effort to date has been expended to identify and correct avoidable data base errors. Many errors, to be expected in the tedious procedure of recording and keypunching vast amounts of data, can be eliminated if sufficient resources are devoted to accomplish the above system of recording and management.

3. SOLAR FLARE FORECASTING METHODS

Solar flare forecasts made by the SESC predict both the occurrence and magnitude of flares. Four classes, denoting the largest event in a 24-hour period can be identified:

(1) No Flare
(2) C Class Flare
(3) M Class Flare
(4) X Class Flare .

It is the moderate class M flare and the major class X flare which are of greatest consequence to near-earth environmental disciplines.

If we let Z represent a random variable such that $Z = 0, 1, 2,$ or 3 corresponding to the largest event (i.e., flare) which occurs in the next 24 hours in a selected region, then estimates are provided by the SESC for the following conditional probabilities:

$$\begin{array}{ll} \text{(i)} & \Pr[Z=0|\underline{x}] \\ \text{(ii)} & \Pr[Z\geq 1|\underline{x}] \\ \text{(iii)} & \Pr[Z\geq 2|\underline{x}] \\ \text{(iv)} & \Pr[Z=3|\underline{x}] \end{array} \qquad (3.0.1)$$

where $\underline{x}$ denotes an observed vector of prediction variables associated with the selected region.

Objective prediction of probabilities (i)-(iv), or variants of these has been accomplished with some success by previous investigators employing a variety of methods. Applicable papers include Hirman, et al. (1980) --discriminant analysis; Vecchia, et al. (1980)--discriminant analysis and logistic regression; and, Jakimiec and Wasiucionek (1980)--multiple linear regression. Successful application of an objective technique on-line at the SESC will enable forecasters to attach immediate quantitative significance to the many interrelated variables comprising the inputs to any forecast method.

In section 3.1, we describe the general problem of probability forecasting and review two methods which are consistent with SESC practice and well-suited for on-line solar flare forecasting using a few variables currently monitored at the SESC. We also briefly discuss the inappropriateness of multiple regression analysis for solar flare forecasting within the present framework. Evaluation measures for probability forecasts are reviewed in section 3.2, and we conclude by noting in section 3.4.2 that regression analysis could be a useful approach to relate flare intensity to active region variables if minor revisions in data collection can be effected.

3.1 Probability Forecast Methods

Let Z denote a random variable and $\underline{x}$ a $p \times 1$ vector of observable random variables used as explanatory or predictor variables for Z. We are interested in the probability structure of Z. All information regarding Z given an observed $\underline{x}$ is contained in the conditional probability distribution, $F(Z|\underline{x}) = Pr[Z \leq z|\underline{x}]$. We therefore define a probability estimate to be an estimate $\hat{F}(Z|\underline{x})$ of the conditional probability distribution. To describe the statistical methods useful for flare forecasting we consider only estimation for a dichotomous random variable where Z=1 (success) or 0 (failure). For example, Z=1 might represent the occurrence of a C, M or X class flare, and Z=0 the non-occurrence of <u>any</u> class of flare. In subsection 3.1.3 we state models for dependent variables with more than two categories and note references for this extension.

3.1.1 Discriminant Analysis (DA)

The general problem is to relate a categorical, or discrete, dependent variable to one or more predictor variables, which may or may not be categorical. That is, based on one or more measurements $\underline{x}$ we wish to classify an observation (element) into one of two or more categories (populations) on the basis of $\underline{x}$. It can be assumed that each category (population) is characterized by a <u>probability distribution</u> of $\underline{x}$.

Suppose that observations belong to two distinct populations P_0 and P_1, characterized by joint probability density functions $f_0(\underline{x})$ and $f_1(\underline{x})$, respectively. For example, P_1 might denote the set of region-day occasions which produce at least a class C flare in the next 24 hours, and P_0 the set of cases producing no flare. Also, let the population membership of a case j be given by the random variable Z_j, where Z_j=1 if $\underline{x}_j$ has distribution f_1, and Z_j=0 if $\underline{x}_j$ is chosen from f_0. Let the prior probability that an observation belongs to P_1 be $Pr[Z_j=1] = p_1$; hence, $Pr[Z_j=0] = p_0 = 1-p_1$.

It is usually assumed that for P_0 and P_1, the random vector $\underline{x}$ has a multivariate normal distribution with different mean vectors $\underline{\mu}_0$ and $\underline{\mu}_1$, but common covariance matrix Σ. That is, for $\underline{x}$ a $p \times 1$ vector we have

$$f_i(\underline{x}) = [(2\pi)^{p/2}|\Sigma|^{1/2}]^{-1} \exp [-(1/2)(\underline{x} - \underline{\mu}_i)'\Sigma^{-1}(\underline{x} - \underline{\mu}_i)].$$

Then for a given random observation and associated vector $\underline{x}$, it can be shown that

$$\Pr[Z=1|\underline{x}] = \{1 + (p_0/p_1) \exp - [\underline{x} - (1/2)(\underline{\mu}_1 + \underline{\mu}_0)]'\Sigma^{-1} [\underline{\mu}_1 - \underline{\mu}_0]\}^{-1}, \qquad (3.1.1)$$

which is of the form

$$\Pr[Z=1|\underline{x}] = \{1 + \exp [-(\alpha + \underline{\beta}'\underline{x})]\}^{-1} \qquad (3.1.2)$$

where $\underline{\beta}'\underline{x} = \sum_{j=1}^{p} \beta_j x_j$.

To construct a probability estimation technique to classify random observations, we require estimates of α and β. For the formulation of the problem assuming normality of $\underline{x}$, which leads to (3.1.1), it is sufficient to estimate $\underline{\mu}_0$, $\underline{\mu}_1$, Σ, and, if necessary, p_0. Suppose $\{\underline{x}_{0j},\ j=1, \ldots, n_0\}$ and $\{\underline{x}_{1j},\ j=1, \ldots, n_1\}$ are random samples of observations from P_0, P_1, respectively. Thus, we are given a set of cases with known population memberships. Then, to estimate α and $\underline{\beta}$ we use

$$\begin{aligned} \hat{\underline{\beta}} &= S^{-1}(\overline{\underline{x}}_1 - \overline{\underline{x}}_0) \\ \hat{\alpha} &= -\mathrm{Ln}(n_0/n_1) - .5\,(\overline{\underline{x}}_1 + \overline{\underline{x}}_0)'\underline{\beta} \end{aligned} \qquad (3.1.3)$$

where S is the pooled estimator of the common covariance matrix Σ, and $\underline{x}_0$ and $\underline{x}_1$ are sample mean vectors. The estimators (3.1.3) will be called discriminant function estimators (DFE's) of $(\alpha, \underline{\beta})$ and by discriminant analysis (DA) we mean the use of DFE's in (3.1.2) to obtain probability estimates for cases with unknown population membership.

3.1.2 Logistic Regression (LR)

The function (3.1.2) is the logistic response function, or logit, a symmetric sigmoid curve. It appears to be a reasonable model for probability forecasting because, as a monotone, smooth function of $\underline{\beta}'\underline{x}$, $Pr[Z=1|\underline{x}]$ is bounded between 0 and 1 and approaches these values as limits as $x_j \to \pm\infty$ for any j.

In section 3.1.1, the assumption that $\underline{x}$ is normally distributed within each population resulted in probability forecasts of the logistic form. However, many types of underlying assumptions about $\underline{x}$ lead also to a prediction equation of the logistic form. For example, the logistic results if some predictor variables are multivariate normal and others are dichotomous, so that the logistic model is appropriate for more general distributions of $\underline{x}$ in addition to multivariate normal.

We will mean by logistic regression (LR) the procedure by which statistical maximum likelihood estimators (MLE's) of $(\alpha, \underline{\beta})$ are obtained for the logistic regression model (3.1.2). Thus, the LR formulation of the problem assumes that the probability function should have the characteristics of the (nonlinear) logit function, and then approximates the true curve by iteratively estimating $(\alpha, \underline{\beta})$ directly. The interested reader is referred to Goldstein and Dillon (1978) or Bishop, et al. (1975) for a further discussion of these concepts.

3.1.3 Extension to Polychotomous Case

In the context of this study, Z may be thought of as the polychotomous random variable representing the largest flare occurring in a future 24-hour period. The dichotomous logistic model, which is the basis for LR and DA, generalizes easily for estimating the probabilities of k events as a function of one or more explanatory variables.

Suppose that observations belong exclusively to one of k populations $P_0, \ldots, P_{k-1}$, characterized by joint probability density functions $f_0(\underline{x}), \ldots, f_{k-1}(\underline{x})$, where $\underline{x}$ is a $p \times 1$ vector of explanatory variables. Also, let the population membership of an observation be given by a random variable Z, where $Z = i$ if $\underline{x}$ is from $f_i(\underline{x})$. If the probability distributions are multivariate normal with common covariance matrix, but different mean vectors, then the posterior probability that an observation is drawn from P_i takes the form

$$P[Z=i|\underline{y}] = \exp(\underline{y}'\underline{\beta}_i) \cdot \left[\sum_{j=0}^{k-1} \exp(\underline{y}'\underline{\beta}_j)\right]^{-1}, \quad i = 0, \ldots, k-1;$$

where $\underline{y}' = [1, x_1, \ldots, x_p]$, and $\underline{\beta}_j' = [\beta_{j0}, \beta_{j1}, \ldots, \beta_{jp}]$.

The one is annexed to the vector of explanatory variables to allow for a constant in the model.

As in the dichotomous case DFE's are obtained if multivariate normality is assumed, and the estimators of the $\underline{\beta}_j$'s are functions of the sample mean vectors, sample covariance matrix, and prior probabilities. LR involves direct maximum likelihood estimation of the $\underline{\beta}_j$'s. For a detailed discussion of LR in the polychotomous case see Jones (1968) and Jones (1975).

3.1.4 Multiple Linear Regression

Although methods for the analysis of categorical data are well developed and have been discussed in the statistical literature for many years, many data analysts of categorical data continue to use inappropriate methods. In particular, discrete variables are often treated as ordinary variables in regression analysis without regard to assumptions of continuity.

Suppose $(Z_j,\underline{x}_j)$, $j=1,\ldots,n$, is a random sample of observations from $(Z,\underline{x})$, where $\underline{x}$ is our usual vector of explanatory variables and Z is 1 or 0 corresponding to the population from which $\underline{x}$ is drawn. Thus, as before, we have a set of cases with known population memberships. Some of the variables $\underline{x}$ may be categorical and some may vary continuously. In such cases $\Pr[Z=1|\underline{x}]$ may be estimated by linear regression methods. The standard regression model would be

$$Z_j = \underline{x}_j'\underline{\beta} + e_j \ , \quad j=1,\ldots,n \quad , \tag{3.1.4}$$

where e_j denotes a random error term with $E(e_j) = 0$; $Var(e_j) = \sigma^2$; $E(e_j e_k) = 0$ if $j \neq k$. $E(\cdot)$ denotes expected value and $Var(\cdot)$ denotes variance. It follows that

$$\begin{aligned} E(Z|\underline{x}) &= \Pr[Z=1|\underline{x}] \\ &= \underline{x}'\underline{\beta} \ , \end{aligned} \tag{3.1.5}$$

where, for convenience, we have dropped the subscript. Then, given the usual regression estimator $\hat{\underline{\beta}}$, from (3.1.4) we obtain a probability estimate for cases with unknown population membership using

$$\hat{Z} = \hat{\Pr}[Z=1|\underline{x}] = \underline{x}'\hat{\underline{\beta}} \ . \tag{3.1.6}$$

This is similar to the Regression Estimation of Event Probabilities (REEP) procedure of Miller (1964).

Use of (3.1.6) for probability forecasting presents some obvious difficulties. Clearly the estimates are not constrained to be between 0 and 1. Further, if it is reasonable that as $\underline{x}'\underline{\beta}$ increases so also does the chance that Z will be one, then true probability forecast function should generally have the S shape (sigmoid) since it must be nondecreasing and bounded between 0 and 1. To approximate such a curve with a straight line may be reasonable over a portion of the curve, but inadmissible for extreme values of $\underline{x}'\underline{\beta}$.

Finally we note also that, under our assumptions, for given $\underline{x}_j$, Z_j is a Bernoulli random variable (see Mood, et al. (1974)). It follows that $E(Z_j|\underline{x}_j) = \underline{x}_j'\underline{\beta}$ and $\mathrm{Var}(Z_j|\underline{x}_j) = \mathrm{Var}(e_j) = \underline{x}_j'\underline{\beta}(1-\underline{x}_j'\underline{\beta})$. This is a direct violation of the equal variance assumptions for (3.1.4), since $\mathrm{Var}(e_j)$ depends on $\underline{x}_j$. Use of ordinary least squares regression estimators under these conditions will yield imprecise predictions.

Possible modifications have been proposed to eliminate technical difficulties in applying linear regression to categorical data. However, extensive delineation of the problems or solutions for the linear model approach is beyond the scope of this report. Theoretical and empirical details regarding inadmissibility of this method may be found in Nerlove and Press (1973) and Brelsford and Jones (1967).

3.2 Scoring Probability Forecasts

For this report the major intended purpose of scoring forecast methods is to provide a measure of the usefulness of proposed statistical models relative to the subjective forecasts of the SESC. It should be emphasized that a complete evaluation of a proposed technique can result only from a long term comparison to the procedure of the SESC. However, as a preliminary measure of utility, the objective flare forecast techniques will be tested by comparing probability estimates from fitted models to the baseline subjective forecasts. Two scoring procedures discussed by Brelsford and Jones (1967) are considered.

Probability estimates or forecasts may be compared by a loss function, $h(Z, \hat{Z})$. The Brier Score (Brier (1950)) used in meteorology is essentially mean square error. For the general model, it assigns a loss of

$$h(Z_{ij}, \hat{Z}_{ij}) = \sum_{j=1}^{k} (Z_{ij}-\hat{Z}_{ij})^2$$

to the i-th trial (the subscript j denotes the event), where we use $\hat{Z}_{ij}$ to denote an estimate of $\Pr[Z_i=j|\underline{x}_i]$, and where $Z_{ij}=1$ if the i-th trial produced event j and 0 otherwise. Given a set of n trials, the Brier Score for a forecast method M is:

$$\mathrm{Brier}(M) = (1/n) \sum_{i=1}^{n} \sum_{j=1}^{k} (Z_{ij}-\hat{Z}_{ij}^{(M)})^2 \, . \tag{3.2.1}$$

Another natural loss function derived from information theory assigns a loss of $-\log Z_{i(m)}$ to the i-th trial, where m is the event which occurred. For forecast method M, the Information Loss Score is given by:

$$\text{Info}(M) = (1/n) \sum_{i=1}^{n} \log \hat{Z}_{i(m)}^{(M)}. \qquad (3.2.2)$$

Given a finite sample of data, this score is minimized by maximum likelihood estimation of the parameters in the logistic function, i.e., by LR estimators of $(\alpha,\underline{\beta})$ in (3.1.2).

For both scoring measures it is desirable to achieve a minimum score. The Information Loss Score may be preferred, however, because probability estimates are constrained to the range $0<\hat{Z}<1$. A probability prediction of zero is unacceptable if the event occurs, since the loss would be $+\infty$.

Evaluation and comparison of probability estimates from many sources is a lengthy subject. General mathematical definitions of scoring methods with desirable practical properties have been determined (see for example, Murphy and Epstein (1967)). A modification by Sanders (1963) partitions the Brier Score into components to determine finer aspects of skill for a given forecast method. The interested reader is referred to Heckman (1979a) for a discussion of these concepts.

3.3 Discussion

Discriminant analysis, logistic regression, and multiple linear regression are only representative of a larger number of possible objective approaches to the solar flare forecasting problem. Alternative techniques are not discussed in this report because, given the current data collection scheme and desired SESC product, we believe that these three statistical methods are the only easily adaptable solutions to the solar flare forecasting problem as presently formulated. We qualify this notion by commenting that the proper application of any statistical method demands approximate consistency of the sample data with the theoretical foundations (i.e., assumptions) of the technique. Our preliminary analyses suggest, in fact, that to accommodate basic assumptions will require clever data manipulation and/or modification of an objective approach.

To determine reasonable compliance with basic assumptions will require the following areas of concern, which are relevant to one or more of the three methods discussed above:

1. Independence of observations. Detailed examination of spatial and/or time correlation properties of solar flare data is essential, but remains to be done. Possible problems have been largely ignored for the current study.

2. Continuity and/or normality of variables. For example, DFE's are derived assuming normality of the explanatory variables. Practically, we require "approximate" conformance to this assumption so that it is sometimes within reason to model discrete scaled data using "normal theory."

3. Invariance of model parameters in time and space. To follow secular variation in the solar cycle may require adaptive models. This may involve a simple periodic recomputation of model estimates but could require a theoretical modification of a particular statistical method. Also, we have not accounted for possible location differences.

4. Equality of variances or covariances (for explanatory variables) among flare class groups. For example, to account for unequal variances could require the inclusion of quadratic terms in a logistic model. We have chosen, instead, to keep the number of variables to a minimum by using variance stabilizing transformations on some variables.

3.4 Recommendations

3.4.1 Probability Forecasting

Press and Wilson (1978) note that:

"Discriminant function estimators have often been used in logistic regression, in both theory and applications. When such estimators were compared empirically with maximum likelihood estimators for logistic regression problems, however, they were found to be generally inferior, although not always by substantial amounts...

The rationale for a logistic formulation of the relationship between qualitative and other variables ... has been discussed extensively in the literature ..."

We suggest consideration of the logistic formulation for the solar flare probability forecasting problem for many of the reasons alluded to in the above statement. For the normal theory case, we derived a logistic response function model and noted that DA is appropriate to obtain a fitted equation for probability forecasting. It was observed, however, that many types of underlying assumptions about explanatory variables lead also to a logistic form of the prediction equation. For example, the logistic results if some predictor variables are multivariate normal and others are dichotomous, so that LR is appropriate for more general distributions than multivariate normal. The DA approach is strictly applicable only if predictors are normally distributed, with complete equality of underlying covariance matrices

To summarize common objections to the general use of DFE's we list the following arguments stated by Press and Wilson (1978):

1. If explanatory variables do not follow a multivariate normal distribution with equal covariance matrices among groups, DFE's of the slope parameters (β_i's) in the logistic function are not "consistent." In particular, this means that if the predictor variables are dichotomous, we cannot expect to obtain accurate forecasts with DFE's, even with an infinite amount of data. This result is proven in Halperin, et al. (1967).

2. When the normality assumption is violated, meaningless variables will tend to be erroneously included in the logistic function with DFE's.

3. Use of DFE's tends to mask troublesome situations. For example, parameter estimates may (correctly) fail to exist with LR, but DFE's may be erroneously computed.

4. There is evidence that DFE's may tend to generate bias in some applications.

For the solar flare probability forecasting problem, most of the explanatory variables are categorical and some are dichotomous. Our present judgment is that LR is a more defensible approach under these circumstances.

We do not provide additional support for the LR approach here. Important comparisons may be found in Brelsford and Jones (1967), Halperin et al. (1967), and Press and Wilson (1978). We remark that in a similar application, LR is used by the National Weather Service to forecast conditional probabilities of frozen precipitation (Glahn, et al. (1973)).

3.4.2 Estimation of Flare Intensity

In section 3.1.4, we reviewed specific objections to the use of multiple linear regression to obtain probability forecasts. The major reason for rejecting linear regression in this study is the inappropriateness of the method to "predict" a discrete dependent variable. Because the choice of forecast methods is dictated, in part, by the chosen scale for recording data, inadmissibility of linear regression is contingent on the present data collection scheme.

Specifically, the coded SESC flare classifications (i.e., classes C, M, X) are actually a categorization of peak X-ray yield, as measured by satellites. Recoding a continuous measurement into a discrete classification may induce a substantial loss of information and greatly restricts the range of valid data analysis methods. Information lost in recoding can be recovered only at great expense, yet, if peak flux were recorded on the original scale, the SESC classes can be assigned with little effort. We suggest that any continuous measurement be recorded in its original scale.

Availability of peak flux measurements could allow the SESC to formulate a linear regression model to forecast actual X-ray yields for active regions. If this were a desirable product, and assuming that the explanatory variables are suitably correlated with flare intensity, linear regression has the following advantages.

1. Linear regression is simple, both conceptually and computationally.

2. Transformed or interaction variables do not increase the complexity of linear regression. Thus, functions of explanatory variables are easily included in the analysis.

3. Confidence intervals for estimated peak flux are readily obtainable.

4. The extension to forecasts for any lead time is straightforward. In particular, if variables were recorded on time scales less than 24 hours, no significant complications arise.

In conclusion, we remark that if explanatory variables are rescaled (presumably yielding "more continuous" scales) during a conversion to less subjective, automatic measurement techniques, desirability of linear and nonlinear regression methods for modeling the solar flare process is likely to increase.

4. SELECTION OF PREDICTORS

4.1 Transformation of Variables

Explanatory or predictor variables may actually be transformations of more basic variables. Because a thorough examination of potential explanatory information may enhance understanding of the solar flare process, the following types of computed variables are identified:

1. Log and power transformations of basic variables. Typical purposes are to stabilize variance or to induce normality.

2. "Functions" of basic variables. Loosely speaking, these represent interactions among variables and should be suggested by expert scientists. Rarely should functions be selected based on empirical evidence of association with flare occurrence.

3. Lagged or rate of change variables. For example, use of lagged variables can account for time correlation in data.

4. Reordered or recategorized basic variables. Some variables may require reordered scales and some may provide as much information with fewer categories.

Careful consideration of transformed variables represents a very large effort in itself. To date, time and resources have not permitted exploration of most of these topics, except to consider log transformations on some variables for stablizing variances among flare class groups. Additionally, to account for suspected interactions involving flare persistence, we chose in our analysis to fit distinct models, conditional on the class of flare occurring during the 24 hour period before the forecast is made. Partitioning of the data into separate segments for analysis avoids introducing covariance (interaction) terms when a variable is known or thought to affect the levels of other variables (see Bishop, et al. (1975), page 359).

4.2 Selection of Basic Variables

Since unavailability of data must be considered in any practical real-time scheme, potential explanatory variables have been divided into three sets based on the observed frequency that data are missing. In Table 3 we display this grouping of variables and the number of cases for which the values are available. The final entry in each column is the total number of cases in each of the four partitions of the data base. Because we have conditioned on the largest flare in the past 24 hours, cases with AGE=1 have been discarded.

A largely heuristic approach was used to select a reduced set of explanatory variables from white light, H-alpha, and historical measurements. It was decided to rank the variables according to some measure of association with variable number 39 (largest flare in the next 24 hours). Variable 46 (largest flare today; persistence) was used to partition the data because of the argument stated in subsection 4.1. This procedure yields four separate rankings of the variables. On this basis we have determined nine variables to be employed as primary forecast criteria in this report. Variables not selected by this ad hoc procedure may prove useful in the future if, for example, interaction variables are allowed. The measure of association used to select variables is described below.

First, for each of the four data segments one-way analysis of variance (AOV) F-ratios were computed for every explanatory variable with variable 39 as the grouping variable. Clearly, if "on the average" the level of a given measurement varies with the event to occur in the next 24 hours, the variable in question may be useful to predict the event.

Because many of the explanatory variables will not satisfy basic assumptions, we do not apply the usual interpretation of the F-ratio or significance levels. In this case we use the statistic represented by an F-ratio to determine <u>relative</u> degrees of utility among explanatory variables, since the "F" statistic has general mathematical properties useful to study differences among sample means.

Table 3. Number of Non-Missing Observations

Variable	Largest Flare Past 24 Hours N	C	M	X
MAGCLAS	2985	574	159	21
SSDYNAM	2968	570	158	21
SSINTER	2982	574	157	21
STGDEV	2963	573	157	21
SECTEOW	2992	574	161	21
NEUTLOR	2953	563	159	20
REVPOL	2992	574	162	21
BRTPTS	2970	571	158	21
PLAGFLUX	2970	571	158	21
ISOPOLE	2972	570	157	21
EFR	2972	570	157	21
AFS	2972	570	157	21
FIRSTAPP	2992	574	162	21
FLUX	2992	574	162	21
RECSPOT	2985	572	159	21
PLAGFIL	2698	537	141	20
NEUTLCOM	2612	524	136	18
NEUTLCHG	2492	498	128	18
ASSOCFIL	2492	502	124	18
RV	1385	312	84	16
MAGSTR	1379	307	84	16
MAGGRAD	1595	296	81	14
	2992	574	162	21

The F statistic chosen to select variables corresponds to the F-ratio used to test for linear trend in the one-way AOV. Let F_N, F_C, F_M, and F_X represent F statistics for a given variable. The subscript denotes the value of the conditioning variable, viz, the event occurring in the past 24 hours. Then the ad hoc measure of association for the variable is represented by the weighted score

$$R = \frac{W_N F_N + W_C F_C + W_M F_M + W_X F_X}{W_N + W_C + W_M + W_X} ,$$

where $W_N = 1$; $W_C = 2$; $W_M = 3$; $W_X = 4$. The weights have been chosen arbitrarily to give greater importance to variables if they are useful to predict flare types following previous flares. This simply acknowledges the fact that most major flares tend to follow other flares.

It must be emphasized that our variable selection procedure is essentially heuristic. With this caution in mind, scores(R) and individual F statistics are presented in Table 4. To assure that F statistics can be compared within groups of variables, cases have been omitted from the analysis if the value of any variable within the particular set is missing. Thus scores or F statistics can be compared within groups, because all F-ratios are based on the same number of degrees of freedom. A relatively high F value is indicative of a stronger degree of association to flare occurrence. Scores have been truncated to the integer part of the computed value.

Based on variable scores from Table 4, we have selected nine variables in three groups to be used as forecast criteria for the models fitted in section 5 of this report. For the nine explanatory variables sample means and standard deviations preceding each largest flare event are shown in Tables 5-7. The variables have been ranked within groups according to their weighted F statistic score. Additionally, statistics have been computed from the same cases used to obtain F values and sample sizes for each group of variables are given at the end of the respective tables. In each table the first block of entries for a variable are sample means based on the number of observations shown at the bottom of the table.

Table 4. Linear Trend "F" Statistics

		Largest Flare Past 24 Hours			
Variable	Score(R)	N	C	M	X
MAGCLAS	52	150.8	86.1	56.3	8.4
SSDYNAM	5	21.8	4.8	2.6	.5
SSINTER	1	.3	2.5	1.1	.9
STGDEV	3	16.8	2.7	.6	1.9
SECTEOW	0	5.3	.1	.0	.0
NEUTLOR	3	22.7	7.8	.1	.0
REVPOL	1	7.3	1.0	.0	.8
BRTPTS	13	78.5	21.3	3.4	.0
PLAGFLUX	10	83.7	.4	.3	4.6
IOSPOLE	5	12.7	5.9	7.9	2.1
EFR	0	.8	1.7	.4	-
AFS	7	44.2	9.5	1.9	.9
FIRSTAPP	3	4.6	3.7	6.1	.8
FLUX	1	9.2	1.2	.9	.6
RECSPOT	47	227.9	71.2	30.8	3.0
PLAGFIL	8	40.8	11.2	6.3	.2
NEUTLCOM	22	111.0	31.4	15.7	.0
NEUTLCHG	4	27.4	5.2	2.1	.4
ASSOCFIL	1	8.0	1.6	.1	.3
RV	2	.8	.4	.8	6.1
MAGSTR	11	18.7	32.1	7.3	2.1
MAGGRAD	18	45.5	29.9	24.8	.2

Table 5. Means and Standard Deviations -- Set 1

Variable	Event Next 24 Hours	Largest Event Past 24 Hours N	C	M	X
MAGCLAS	N	1.71	2.22	2.32	2.50
	C	2.18	2.90	3.62	4.60
	M	2.20	3.62	4.13	5.42
	X	4.50	3.25	5.50	5.83
	N	.61	.81	.83	.70
	C	.91	1.37	1.60	1.81
	M	.71	1.54	1.58	1.27
	X	2.12	1.75	1.00	1.16
RECSPOT	N	1.41	2.46	3.08	5.00
	C	2.36	3.53	4.60	5.20
	M	2.97	4.62	5.58	7.28
	X	2.00	4.87	6.00	6.83
	N	.98	1.90	2.14	4.24
	C	1.77	2.14	2.47	2.77
	M	2.20	2.48	2.32	.75
	X	0	2.03	1.82	.98
BRTPTS	N	.24	.67	.87	1.50
	C	.61	1.06	1.38	1.00
	M	.51	1.07	1.09	1.42
	X	.50	.75	1.50	1.16
	N	.57	.79	.83	.70
	C	.79	.82	.75	.70
	M	.74	.84	.78	.78
	X	.70	.88	.57	.75
PLAGFLUX	N	.12	.33	.42	.50
	C	.26	.35	.34	.60
	M	.48	.35	.34	.28
	X	1.00	.50	.50	0
	N	.32	.47	.49	.70
	C	.44	.47	.47	.54
	M	.50	.48	.48	.48
	X	0	.53	.57	0
AFS	N	.12	.20	.21	.50
	C	.28	.35	.30	.40
	M	.22	.33	.11	.28
	X	0	.25	0	.16
	N	.32	.40	.41	.70
	C	.45	.48	.46	.54
	M	.42	.47	.32	.48
	X	0	.46	0	.40
Number of Cases	N	2583	304	56	2
	C	261	190	50	5
	M	35	51	43	7
	X	2	8	4	6

Table 6. Means and Standard Deviations -- Set 2

Variable	Event Next 24 Hours	Largest Event Past 24 Hours N	C	M	X
NEUTLCOM	N	.84	1.40	1.58	4.00
	C	1.38	1.83	2.02	2.40
	M	1.52	1.97	2.61	2.50
	X	2.50	2.37	2.50	3.40
	N	.77	.90	1.11	0
	C	.98	.95	.98	1.51
	M	1.03	.97	1.17	.83
	X	.70	1.30	.57	.54
PLAGFIL	N	.99	1.98	2.17	2.50
	C	1.57	2.42	3.19	2.60
	M	1.47	2.88	2.74	3.66
	X	2.50	1.87	3.75	2.80
	N	1.29	1.60	1.46	.70
	C	1.49	1.52	1.24	1.81
	M	1.37	1.46	1.31	1.50
	X	3.53	.99	.95	1.09
Number of Cases	N	2108	250	41	2
	C	228	177	41	5
	M	23	42	31	6
	X	2	8	4	5

Table 7. Means and Standard Deviations -- Set 3

Variable	Event Next 24 Hours	Largest Event Past 24 Hours N	C	M	X
MAGGRAD	N	.04	.08	.06	.10
	C	.08	.12	.17	.21
	M	.06	.17	.16	.32
	X	.15	.15	.27	.20
	N	.05	.08	.06	0
	C	.06	.08	.09	.13
	M	.05	.09	.08	.17
	X	.14	.10	.08	.06
MAGSTR	N	15.38	16.55	17.26	15.00
	C	17.37	18.16	19.44	21.00
	M	17.43	20.35	20.04	21.60
	X	18.50	20.66	22.50	22.50
	N	4.61	3.84	3.97	0
	C	4.81	3.59	4.52	4.83
	M	4.42	3.27	3.68	3.04
	X	2.12	4.63	2.12	3.10
Number of Cases	N	1009	143	26	1
	C	104	98	25	4
	M	16	28	23	5
	X	2	6	2	4

4.3 Variable Rescaling

From the information provided in Tables 5-7, it was decided to transform some of the nine explanatory variables to a log scale. The decision to rescale any given variable was based on an (intuitive) examination of sample standard deviations within data segments. It is our purpose in this case to avoid the introduction of quadratic terms into the analysis, since these can be shown to be necessary in the logistic function if covariance matrices are not equal among groups. Again our decisions are heuristic and should be reconsidered in any continuation of this work.

The nine selected forecast criteria, in rescaled form, are listed below. Log transformations are base 10.

$$
\begin{array}{llll}
x_1 = \text{Log} & (\text{MAGCLAS} & + & .5) \\
x_2 = \text{Log} & (\text{RECSPOT} & + & .5) \\
x_3 = \text{Log} & (\text{BRTPTS} & + & 1.5) \\
x_4 = & \text{PLAGFLUX} & + & 1.0 \\
x_5 = & \text{AFS} & + & 1.0 \\
\hline
x_6 = \text{Log} & (\text{NEUTLCOM} & + & 1.5) \\
x_7 = \text{Log} & (\text{PLAGFIL} & + & 1.5) \\
\hline
x_8 = \text{Log} & (\text{MAGGRAD} & + & 1.5) \\
x_9 = & \text{MAGSTR} & + & 1.0
\end{array}
\qquad (4.3.1)
$$

In section 5, probability forecasts are obtained for both LR and DA using these nine explanatory variables.

5. EXAMPLE

To illustrate the ideas of section 3, we consider estimation of the following conditional probabilities:

$$
\begin{array}{ll}
\text{(i)} & \Pr[Z=0|\underline{x}] \\
\text{(ii)} & \Pr[Z \geq 1|\underline{x}] \\
\text{(iii)} & \Pr[Z \geq 2|\underline{x}] \\
\text{(iv)} & \Pr[Z=3|\underline{x}]
\end{array}
\quad , \qquad (5.0.1)
$$

where Z = 0, 1, 2, or 3 corresponding to the largest flare which occurs in the next 24 hours in a selected region. Four separate models, identified by subsets of the nine variables in (4.3.1), are considered:

$$
\begin{array}{ll}
\text{Model I:} & \{x_1, \ldots, x_5\} \\
\text{Model II:} & \{x_1, \ldots, x_7\} \\
\text{Model III:} & \{x_1, \ldots, x_9\} \\
\text{Model IV:} & \{x_1, x_4\}
\end{array}
\quad . \qquad (5.0.2)
$$

Because we have decided to condition on flare persistence, fitting distinct models following each flare class, Model IV is included to forecast cases following class X flare since the low number of available cases dictates that only a few parameters be estimated. Each of models I-III is fitted conditional on No, C, or M class flare occurrence. The combinations of multiple models and conditioning on persistence gives rise to the models

indicated by an "*" in Table 8. Both LR and DA lead to fitted models which are well suited for on-line forecasting. For the examples, the methods are compared to the baseline SESC forecast.

Table 8. Forecast Models

	Largest Event Past 24 Hours			
Model	No	C	M	X
I	*	*	*	
II	*	*	*	
III	*	*	*	
IV				*

5.1 Case Selection

The current analysis utilizes on 6097 region-day records collected from January 1, 1977 to January 31, 1979. These records were first reduced to 4487 records by the elimination of records indicating the absence of sunspots, since such regions rarely produce flares. For each of the conditional models in Table 8, parameters are estimated using all cases with non-missing data on variables associated with the particular model number, according to the lists in (4.3.2). That is, fitted models are not based on a common set of cases. Additionally, because the number of cases following the no flare event exceeded computer program limitations, models for column one in Table 8 are estimated using cases after December 1977.

5.2 Grouping of Cases

As k increases, the polychotomous logistic model requires increasing numbers of parameters to be estimated. With the present data base and approach, some of the models must be estimated based on a relatively small number of available cases. Because the estimation of a large number of parameters is subject to criticism under these circumstances, it was decided to use a dichotomous (k=2) model at each stage of parameter estimation. This is accomplished by regrouping and recoding cases depending on which probability in (5.0.1) is being considered. Each combination of model-persistence-event leads to one set of estimated parameters, avoiding estimation of a 4-category model. According to this scheme, Table 9 displays the manner in which cases are combined for each model-persistence combination. Note that $Pr[Z=0|\underline{x}]$ is not listed but can be obtained by subtraction.

Table 9. Case Grouping

Probability	Case Groups	
$Pr[Z>1|\underline{x}]$	N	vs. C, M, X
$Pr[Z>2|\underline{x}]$	N, C	vs. M, X
$Pr[Z=3|\underline{x}]$	N, C, M	vs. X

We conclude this section by remarking that the (unorthodox) scheme of recoding dependent variable cases into a dichotomy at each stage but reconstructing a 4-category type analysis will generally have disturbing theoretical implications. Two such problems are: (1) probabilities may no longer be additive when obtained by subtraction. That is, $\sum \Pr[Z=j|\underline{x}]$ may not be 1; and, (2) if the equal covariance matrix assumption is satisfied in the 4-category case, it may necessarily be violated by constructing a dichotomy. For these reasons, the present approach should be viewed as ad hoc, and should be thoroughly evaluated if continued. It is likely, however, that if data management problems are corrected, the resulting increase in reliable data will allow use of the 4-category model, thus eliminating the need for regrouping of cases.

5.3 Validation

In most analyses of the logistic type, it is useful to divide the data cases into two groups --a training set and a validation set. The validation set is held out of the parameter estimation phase of the analysis, but used later to cross-validate the probability forecast function estimated from the training set. Normally, if the data are unordered, the two samples are obtained by completely random subdivision of the cases. Considering the sequential nature of the data in the current application, a more logical approach is to hold out data after a selected date to be used for validation "into the future." A variation of this idea was used by Hirman, et al. (1980) using a sliding data window.

We agree that time ordered validation is the sensible approach and is naturally consistent with real-time evaluation of forecast methods. However, small samples generated by the conditional model approach have precluded validation for the examples in this report. We have chosen to use all data to obtain parameter estimates which are, hopefully, more accurate and precise. Model validation (or invalidation) can be easily accomplished using the "cleaned" data set for February 1, 1979 to June 10, 1979 when it becomes available.

5.4 Parameter Estimation

In this section we obtain LR estimators of $(\alpha, \underline{\beta})$ for each model-persistence-event combination. Corresponding DA estimators are not listed, but have been used for comparison in later sections.

Let the flare events to be forecast be given by

$$E_1 = \{Z \geq 1\}$$
$$E_2 = \{Z \geq 2\}$$
$$E_3 = \{Z = 3\} \quad ,$$

and let $(\hat{\alpha},\hat{\underline{\beta}})_{mpk}$ represent estimators of $(\alpha,\underline{\beta})$ for model m (i.e., I, II, III, or IV), persistence value p (i.e., N, C, M or X), and event E_k. Then the probability forecast on case j for event E_k is

$$\hat{Pr}[E_k|\underline{x}_j] = \{1 + \exp - (\hat{\alpha} + \hat{\underline{\beta}}'\underline{x}_j)\}^{-1} \quad , \tag{5.4.1}$$

where, for convenience, we have dropped subscripts on the estimators $(\hat{\alpha},\hat{\underline{\beta}})$. Here $\underline{x}_j$ denotes the set of observed variables on case j corresponding to the appropriate model. For example, with model I, $\underline{\beta}'\underline{x}_j = \beta_1 x_{1j} + \cdots + \beta_5 x_{5j}$. Recall that DA or LR forecasts derive from (5.4.1) depending on the estimation method used to obtain $(\hat{\alpha},\hat{\underline{\beta}})$.

To determine the usefulness of explanatory variables to predict flares, a chi-square test of significance was computed for each model-persistence-event triplet. Entries in Table 10 are significance levels which have been used to identify circumstances where explanatory variables (apparently) do not provide information associated with flare occurrence.

Table 10. Overall Chi-Square Significance Levels[1]

Model	Largest Event Past 24 Hours	Flare Event to be Forecast: C, M or X	M or X	X
I	No Flare	.0000	.0000	-
	C Flare	.0000	.0000	.3735
	M Flare	.0000	.0000	.0387
II	No Flare	.0000	.0000	-
	C Flare	.0000	.0000	.2772
	M Flare	.0000	.0002	.1109
III	No Flare	.0000	.0075	-
	C Flare	.0000	.0000	.2144
	M Flare	.0000	.0004	-
IV	X Flare	.0012	.0115	.0339

[1]Values are not reported if the data are insufficient for stable estimation of parameters.

Tables 11-14 display estimated parameters for models with significance levels less than .20. Individual underlined coefficients are those significantly non-zero (at the 5% significance level), but conditional on all other variables already in the logistic function. Estimated coefficients correspond to transformed variables though, for brevity, we have identified parameters with the original variable names. That is, β_j corresponds to x_j where $(x_1,\cdots,x_9)$ are defined in (4.3.1).

Table 11. Estimated Parameters—Model I

Largest Event Past 24 Hours[1]	Variable	Parameter Estimated	Flare Event to be Forecast C, M or X	M or X	X
No Flare	Constant	α	- 5.228	- 7.487	
•2089 N	MAGCLAS	β_1	$\underline{2.852}$	2.162	
• 226 C	RECSPOT	β_2	$\underline{1.738}$	$\underline{1.586}$	
• 32 M	BRTPTS	β_3	$\underline{1.813}$	1.418	
• 1 X	PLAGFLUX	β_4	$\underline{.423}$	$\underline{1.249}$	
	AFS	β_5	$\underline{.447}$	- .133	
C Flare	Constant	α	- 3.600	- 4.656	
• 309 N	MAGCLAS	β_1	$\underline{4.092}$	$\underline{3.898}$	
• 199 C	RECSPOT	β_2	$\underline{.999}$	1.104	
• 54 M	BRTPTS	β_3	$\underline{1.678}$	- .034	
• 8 X	PLAGFLUX	β_4	- .065	- .049	
	AFS	β_5	.428	.037	
M Flare	Constant	α	- 4.463	- 3.050	- 5.284
• 56 N	MAGCLAS	β_1	$\underline{8.368}$	$\underline{4.024}$	10.357
• 51 C	RECSPOT	β_2	.407	$\underline{2.158}$	- .209
• 45 M	BRTPTS	β_3	2.704	- 1.122	5.743
• 4 X	PLAGFLUX	β_4	- .330	.098	1.247
	AFS	β_5	- .086	- .989	- 9.226

[1] Also indicated are sample sizes for flare types based on Variable 39.

Table 12. Estimated Parameters--Model II

Largest Event Past 24 Hours	Variable	Parameter Estimated	Flare Event to be Forecast C, M or X	M or X	X
No Flare	Constant	α	- 5.437	- 7.955	
•1761 N	MAGCLAS	β_1	2.724	2.134	
• 206 C	RECSPOT	β_2	1.412	1.510	
• 26 M	BRTPTS	β_3	1.047	1.619	
• 1 X	PLAGFLUX	β_4	.270	1.213	
	AFS	β_5	.359	- .344	
	NEUTLCOM	β_6	1.984	1.610	
	PLAGFIL	β_7	.279	- .000	
C Flare	Constant	α	- 3.860	- 5.140	
• 270 N	MAGCLAS	β_1	3.862	3.638	
• 186 C	RECSPOT	β_2	.928	1.053	
• 47 M	BRTPTS	β_3	1.425	.326	
• 8 X	PLAGFLUX	β_4	- .157	- .037	
	AFS	β_5	.500	- .009	
	NEUTLCOM	β_6	.838	.315	
	PLAGFIL	β_7	.279	.638	
M Flare	Constant	α	- 6.056	- 3.635	
• 47 N	MAGCLAS	β_1	7.876	3.508	
• 44 C	RECSPOT	β_2	.396	2.232	
• 38 M	BRTPTS	β_3	2.853	- .209	
• 4 X	PLAGFLUX	β_4	- .019	.286	
	AFS	β_5	- .295	- .904	
	NEUTLCOM	β_6	.310	1.375	
	PLAGFIL	β_7	2.442	- 1.161	

Table 13. Estimated Parameters--Model III

Largest Event Past 24 Hours	Variable	Parameter Estimated	Flare Event to be Forecast C, M or X	M or X	X
No Flare	Constant	α	- 8.791	- 6.678	
• 557 N	MAGCLAS	β_1	.523	.574	
• 71 C	RECSPOT	β_2	.779	1.759	
• 12 M	BRTPTS	β_3	1.171	2.337	
• 1 X	PLAGFLUX	β_4	.758	1.439	
	AFS	β_5	.518	- .530	
	NEUTLCOM	β_6	2.782	3.340	
	PLAGFIL	β_7	.819	.185	
	MAGGRAD	β_8	9.749	-16.455	
	MAGSTR	β_9	.065	.085	
C Flare	Constant	α	- 5.574	- 8.465	
• 132 N	MAGCLAS	β_1	4.132	3.060	
• 89 C	RECSPOT	β_2	.672	1.988	
• 24 M	BRTPTS	β_3	.934	.350	
• 6 X	PLAGFLUX	β_4	.110	.184	
	AFS	β_5	.392	- .379	
	NEUTLCOM	β_6	.856	- .055	
	PLAGFIL	β_7	.125	.120	
	MAGGRAD	β_8	- .756	.427	
	MAGSTR	β_9	.100	1.815	
M Flare	Constant	α	-11.307	- 5.371	
• 25 N	MAGCLAS	β_1	10.884	2.750	
• 22 C	RECSPOT	β_2	- .629	3.666	
• 22 M	BRTPTS	β_3	1.365	- 1.663	
• 2 X	PLAGFLUX	β_4	- .932	- .812	
	AFS	β_5	- .592	- .898	
	NEUTLCOM	β_6	2.266	4.583	
	PLAGFIL	β_7	.474	- 4.674	
	MAGGRAD	β_8	37.698	16.784	
	MAGSTR	β_9	.057	- .010	

Table 14. Estimated Parameters--Model IV

Largest Event Past 24 Hours	Variable	Parameter Estimated	Flare Event to be Forecast C, M or X	M or X	X
X Flare	Constant	α	- 43.571	- 3.560	4.709
• 3 N					
• 5 C	MAGCLAS	β_1	105.350	8.554	5.350
• 7 M	PLAGFLUX	β_4	- 6.311	- 1.416	- 9.131
• 6 X					

Associations of the explanatory variables with flare incidence may be (roughly) studied by inspection of the estimated prediction equations. Estimates of β_j's which are positive are indicative of positive association of the corresponding x_j to flare occurrence. Note, for example, that (very roughly speaking) plage fluctuations seem to be either positively associated or unassociated with the occurrence of flares except following the occurrence of X flares, when the absence of plage fluctuations may be associated with persistence of large flares. We emphasize that caution must be exercised in making statements like the one above, which was presented to illustrate the interpretation of model coefficients. Clearly, it is possible that apparent associations are purely spurious, so we should take great care to interpret results.

Because explanatory variables have different scales of measurement, it is not possible to interpret directly the magnitude of estimated parameters in Tables 11-14. A procedure to facilitate interpretation of parameters by computing standardized coefficients is available and should be considered in any continuation of this work. The reader is referred to Bishop, et al. (1975) for a discussion of this technique.

5.5 Model Evaluation

We are particularly interested in a comparison of actual predictions by the objective and subjective procedures. In this section Brier scores for SESC, DA, and LR forecasts are presented. Recall that, at this time, we do not have reliable data for cross-validation of results. We remark here that it is reasonable to suspect that some cases for which large discrepancies between SESC and objective methods exist could be the result of data tabulation or keypunch error. This would not, in principle, be a problem if suggestions for data collection and management could be successfully implemented (see section 2). Existence of outliers will disrupt the estimation of parameters and evaluation of models.

Probability estimates were obtained for all observations using estimated logit functions for both DA and LR. Detailed classification tables for the cases are given in Appendix B. Because Brier (or Information Loss) functions indicate the "average" discrepancy between the probability estimates for events and a posteriori probabilities (viz, 1 for the event which occurred, 0 for other events), scores are more informative measures of forecast method performance than classification tables. Brier scores are given in Tables 15-17.

Tables 15-17 are intended, primarily, to illustrate comparison of probability forecast methods. Emphasizing that LR and DA scores are based on the same data used to estimate parameters, we note (with caution) that, where applicable, LR generally has lower total scores than SESC or DA, but differences with DA are nearly negligible. The latter result was expected since the log transformation tends to induce normality. Detailed analysis of scores will not be presented in the absence of a validation data set.

In conclusion, we remark that to facilitate interpretation of Brier scores in Tables 15-17, one may compute the square root of (B/2), where B is the table entry. The result can be thought of as the "average deviation of the probability estimate from 1 for the event which occurred." For example, a Brier score of .25 corresponds an "average probability deviation" of .35.

Table 15. Brier Scores for Pr[C, M or X Flare | x]

Model	Largest Event Past 24 Hours	Event Observed	Number Cases	Forecast Source SESC	LR	DA
I	No Flare	No Flare	2089	.200	.034	.048
		C Flare	226	.839	1.361	1.294
		M Flare	32	.737	1.268	1.158
		X Flare	1	.020	1.053	.971
		All	2348	.269	.179	.183
	C Flare	No Flare	309	.681	.351	.347
		C Flare	199	.319	.516	.520
		M Flare	54	.167	.403	.410
		X Flare	8	.178	.457	.457
		All	570	.499	.415	.415
	M Flare	No Flare	56	1.156	.505	.501
		C Flare	51	.112	.296	.301
		M Flare	45	.031	.171	.173
		X Flare	4	.026	.006	.008
		All	156	.461	.328	.328
II	No Flare	No Flare	1761	.204	.038	.050
		C Flare	206	.832	1.327	1.273
		M Flare	26	.700	1.211	1.103
		X Flare	1	.020	1.045	.997
		All	1994	.275	.187	.191
	C Flare	No Flare	270	.691	.367	.363
		C Flare	186	.325	.488	.493
		M Flare	47	.174	.372	.379
		X Flare	8	.178	.434	.430
		All	511	.502	.412	.413
	M Flare	No Flare	47	1.137	.487	.485
		C Flare	44	.077	.243	.246
		M Flare	38	.035	.188	.191
		X Flare	4	.026	.004	.005
		All	133	.438	.306	.308
III	No Flare	No Flare	557	.245	.051	.060
		C Flare	71	.742	1.208	1.174
		M Flare	12	.733	1.066	.994
		X Flare	1	.020	1.073	.989
		All	641	.309	.200	.202
	C Flare	No Flare	132	.737	.364	.360
		C Flare	89	.361	.499	.505
		M Flare	24	.178	.254	.256
		X Flare	6	.230	.412	.404
		All	251	.538	.402	.402
	M Flare	No Flare	25	.941	.327	.286
		C Flare	22	.062	.236	.261
		M Flare	22	.028	.151	.184
		X Flare	2	.013	.000	.000
		All	71	.360	.235	.239
IV	X Flare	No Flare	3	1.433	.296	.426
		C Flare	5	.013	.044	.152
		M Flare	7	.109	.032	.012
		X Flare	6	.010	.000	.001
		All	21	.247	.063	.101

Table 16. Brier Scores for Pr[M or X Flare | x]

Model	Largest Event Past 24 Hours	Event Observed	Number Cases	Forecast Source SESC	LR	DA
I	No Flare	No Flare	2089	.024	.001	.007
		C Flare	226	.085	.003	.020
		M Flare	32	1.518	1.854	1.683
		X Flare	1	.080	1.767	1.672
		All	2348	.050	.027	.032
	C Flare	No Flare	309	.126	.019	.021
		C Flare	199	.270	.055	.067
		M Flare	54	.919	1.384	1.350
		X Flare	8	.768	1.427	1.399
		All	570	.261	.181	.182
	M Flare	No Flare	56	.327	.104	.103
		C Flare	51	.717	.284	.295
		M Flare	45	.343	.704	.697
		X Flare	4	.420	.344	.313
		All	156	.461	.342	.343
II	No Flare	No Flare	1761	.024	.001	.007
		C Flare	206	.088	.003	.020
		M Flare	26	1.514	1.842	1.640
		X Flare	1	.080	1.733	1.629
		All	1994	.050	.026	.031
	C Flare	No Flare	270	.121	.018	.019
		C Flare	186	.275	.053	.065
		M Flare	47	.911	1.366	1.333
		X Flare	8	.768	1.455	1.433
		All	511	.260	.177	.179
	M Flare	No Flare	47	.294	.112	.112
		C Flare	44	.701	.276	.282
		M Flare	38	.368	.706	.702
		X Flare	4	.420	.390	.357
		All	133	.454	.344	.344
III	No Flare	No Flare	557	.034	.002	.010
		C Flare	71	.081	.006	.015
		M Flare	12	1.521	1.730	1.581
		X Flare	1	.080	1.599	1.446
		All	641	.067	.038	.042
	C Flare	No Flare	132	.117	.022	.026
		C Flare	89	.262	.074	.084
		M Flare	24	.845	1.076	1.040
		X Flare	6	.710	1.151	1.167
		All	251	.252	.168	.171
	M Flare	No Flare	25	.280	.123	.126
		C Flare	22	.635	.255	.271
		M Flare	22	.357	.511	.510
		X Flare	2	.340	.081	.063
		All	71	.416	.283	.288
IV	X Flare	No Flare	3	.875	.055	.019
		C Flare	5	1.201	.616	.623
		M Flare	7	.112	.282	.291
		X Flare	6	.042	.071	.051
		All	21	.460	.269	.262

Table 17. Brier Scores for Pr[X Flare | x]

Model	Largest Event Past 24 Hours	Event Observed	Number Cases	SESC	Forecast Source LR	DA
I	No Flare	No Flare	2089	.002		
		C Flare	226	.004		
		M Flare	32	.003		
		X Flare	1	1.280		
		All	2348	.003		
	C Flare	No Flare	309	.009		
		C Flare	199	.029		
		M Flare	54	.051		
		X Flare	8	1.568		
		All	570	.042		
	M Flare	No Flare	56	.020	.001	.001
		C Flare	51	.115	.013	.008
		M Flare	45	.142	.008	.006
		X Flare	4	1.462	1.505	1.560
		All	156	.123	.046	.045
II	No Flare	No Flare	1761	.002		
		C Flare	206	.005		
		M Flare	26	.003		
		X Flare	1	1.280		
		All	1994	.003		
	C Flare	No Flare	270	.007		
		C Flare	186	.030		
		M Flare	47	.052		
		X Flare	8	1.568		
		All	511	.044		
	M Flare	No Flare	47	.022		
		C Flare	44	.127		
		M Flare	38	.131		
		X Flare	4	1.462		
		All	133	.131		
III	No Flare	No Flare	557	.002		
		C Flare	71	.005		
		M Flare	12	.004		
		X Flare	1	1.280		
		All	641	.004		
	C Flare	No Flare	132	.008		
		C Flare	89	.029		
		M Flare	24	.065		
		X Flare	6	1.494		
		All	251	.056		
	M Flare	No Flare	25	.024		
		C Flare	22	.104		
		M Flare	22	.129		
		X Flare	2	1.300		
		All	71	.117		
IV	X Flare	No Flare	3	.128	.011	.013
		C Flare	5	.108	.143	.147
		M Flare	7	.297	.309	.311
		X Flare	6	.951	.602	.603
		All	21	.415	.311	.313

6. SUMMARY AND SUGGESTIONS

Even though the difficulties encountered in the course of the study precluded a rigorous analysis of the data, the objective technique of logistic regression has been demonstrated to be potentially useful for probability forecasting of solar flares. This conclusion is based on evidence of statistical association of solar flare incidence with many of the region analysis variables collected by the SESC. Because many procedures for this study were developed heuristically, all conclusions should be evaluated with caution. To summarize steps of the analysis we list important considerations which lead to the examples in section 5:

1. A priori elimination of no sunspot cases and data collected after January 31, 1979.

2. A priori elimination of location variables and a few troublesome region analysis variables.

3. Selection of discriminant analysis and logistic regression as the only methods to be considered for objective probability forecasting.

4. Partitioning the data into four segments (conditional on the class of flare occurring during the past 24 hours) to avoid introducing covariate terms into the analysis.

5. Heuristic selection of nine basic explanatory variables, divided into three sets depending on the frequency of missing values.

6. Rescaling of some of the nine basic variables based on (intuitive) examination of sample standard deviations.

In the examples we considered the prediction of Pr[C, M, or X Flare | $\underline{x}$], Pr[M or X Flare | $\underline{x}$], and Pr[X Flare | $\underline{x}$], where $\underline{x}$ denotes a subset of the nine transformed variables. These three probabilities are those estimated by the SESC.

Developing the full potential of objective forecast techniques will demand great effort and cooperation on a broad front. We have identified the central considerations for continuation of this work to include:

1. Devotion of sufficient resources for data collection, correction and management. No progress can be expected if reliable information is not available (see section 2.1).

2. Re-evaluation of objectives. Are flare intensity predictions desirable? Would forecasts of other related probabilities be useful, e.g., Pr[M Flare | $\underline{x}$]?

3. Consideration of transformed, lagged, rate-of-change, and interaction variables.

4. Stepwise selection of variables.

5. Evaluation of the need for conditional models to account for persistence.

6. Examination of adaptive models to follow secular variation in the solar cycle.

7. Study of spatial and time correlation (possibly accounted for by conditional models or inclusion of lagged or location variables).

The inherent stochastic nature of solar flare phenomena should dictate that statistical methods, particularly in the fields of multivariate analysis and stochastic processes, be developed in the direction of specific peculiarities arising in the flare prediction problem. It should be clearly pointed out that submitting solar flare data to various cookbook methods will not necessarily yield the most efficient analysis. Future determination of a multitude of pertinent features of the data could very likely depend on the development of appropriate theories and procedures based on intuitive leads by solar scientists regarding plausible stochastic models for solar flares.

Perhaps the single most important consideration for further investigations concerns the types of variables to be recorded and their scale and time of measurement. Understanding based on stochastic models relating region analysis data to flare occurrence may not be achieved if the informative variables are not first determined and then properly collected. Such basic problems, if not satisfactorily resolved, may result in future solar flare records which cannot possibly provide the statistical information of interest. Thus, the need for statistical planning in this field cannot be exaggerated.

7. ACKNOWLEDGEMENT

This study was supported jointly by the the U.S. Air Force and the U.S. Department of Commerce, NOAA Space Environment Laboratory, Project No. NA79RAA03613. The authors are indebted to a large number of persons whose expertise and thoughtful discussion contributed to an understanding of the problem. The list below is by no means complete.

NOAA	USAF
T. B. Gray	D. F. Neidig
G. R. Heckman	T. W. Metzger
J. W. Hirman	
W. E. Flowers	
T. G. Clark	
C. M. Cruickshank	

The authors are especially grateful to J. E. Koontz of the NBS Statistical Engineering Division (SED) for the expert development of most of the computer software needed to undertake this study. We also thank J. R. Donaldson of the SED for additional computational assistance and Ms. Lorna Buhse for the typing of the manuscript.

REFERENCES

Bishop, M., Fienberg, S. E., and Holland, P. W. (1975): Discrete Multivariate Analysis: Theory and Practice. Cambridge, Massachusetts: MIT Press.

Brelsford, M. W., and Jones, R. H. (1967): Estimating probabilities. Monthly Weather Review, 95:570-576.

Brier, G. W. (1950): Verification of forecasts expressed in terms of probability. Monthly Weather Review, 78:1.

Glahn, H. R., Brocchieri, J. R., and Jones, R. H. (1973): Forecasting the conditional probability of frozen precipitation. In: Proceedings from the Third Conference on Probability and Statistics in Atmospheric Science, June 19-22, 1973, Boulder, Colorado, American Meteorological Society.

Goldstein, M., and Dillon, W. R. (1978): Discrete Discriminant Analysis. New York: Wiley.

Halperin, M., Blackwelder, W. C., and Vester, J. L. (1971): Estimation of the multivariate logistic risk function: a comparison of discriminant function and maximum likelihood approaches. Journal of Chronic Disease, 24:125-158.

Heckman, G. R. (1979a): Verification of solar flare forecasts at the space environment services center for the years 1969-1974. Technical report (to appear), U.S. Department of Commerce, National Oceanic and Atmospheric Administration.

Heckman, G. R. (1979b): Predictions of the Space Environment Services Center. Solar-Terrestrial Predictions Proceedings, Vol. 1, (R. F. Donnelly, Editor), U.S. Department of Commerce, National Oceanic and Atmospheric Administration, Boulder, Colorado.

Hirman J. W., and Flowers, W. E. (1978): An objective approach to region analysis for flare forecasting. Proceedings of the Symposium on the Effect of the Ionosphere on Space Systems and Communications, Naval Research Laboratory, Washington, DC. (in press).

Hirman, J. W., Neidig, D. F., Seagraves, P. H., Flowers, W. E., and Wiborg, P. H. (1980): The application of multivariate discriminant analysis to solar flare forecasting. Solar-Terrestrial Prediction Proceedings, Vol.3, U.S. Department of Commerce, National Oceanic and Atmospheric Administration, Boulder, Colorado.

Jakimiec, M., and Wasiucionek, J. (1980): Short-term flare predictions and their stationarity during the 11-year cycle. Solar-Terrestrial Predictions Proceedings, Vol. 3, U.S. Department of Commerce, National Oceanic and Atmospheric Administration, Boulder, Colorado.

Jones, R. H. (1968): A nonlinear model for estimating probabilities of k events. Monthly Weather Review, 96:383-384.

Jones, R. H. (1975): Probability estimation using a multinomial logistic function. J. Statis. Comput. Simul., 3:315-329.

Mangis, S. J. (1975): Introduction to solar-terrestrial phenomena and the Space Environment Services Center. NOAA Technical Report ERL 315-SEL32, U.S. Department of Commerce, National Oceanic and Atmospheric Administration, Boulder, Colorado.

Miller, R. G. (1964): Regression estimation of event probabilities. Technical Report, Travelers Research Center.

Mood, A. M., Graybill, F. H., and Boes, D. C. (1974): Introduction to the Theory of Statistics, New York: McGraw-Hill.

Murphy, A. H., and Epstein, E. S. (1967): A note on probability forecasting and "hedging." J. Appl. Meteor., 6:1002.

Nerlove, M., and Press, S. J. (1973): Univariate and multivariate log-linear and logistic models. Rand Corporation Technical Report R-1306-EDA/NIH, Santa Monica, CA.

Nie, N. H., Hull., Jenkins J. G., Steinbrenner, K., and Bent, D. H. (1975): Statistical Package for the Social Sciences, Second Edition, New York: McGraw-Hill.

Press, S. J., and Wilson, S. (1978): Choosing between logistic regression and discriminant analysis. Journal of the American Statistical Association, 73:699-705.

Sanders, F. (1963): On subjective probability forecasting, J. Appl. Meteor., 2:191.

Vecchia, D. F., Caldwell, G. A., Tryon, P. V., and Jones, R. H. (1980): Logistic regression for solar flare probability forecasting. Solar-Terrestrial Predictions Proceedings, Vol. 3, U.S. Department of Commerce, National Oceanic and Atmospheric Administration, Boulder, Colorado.

APPENDIX A

Region Analysis Variables

This analysis is based on data reported or observed during the past 24 hours only (0000 UT to 2400 UT) and should be completed in time for 2200Z forecast.

Region Analysis

LOCATION

1.	DATE	Year, month, day.
2.	REGION	Region number.
3.	APPLONG	First appearance longitude.
4.	CURLONG	Current longitude.
5.	NSLAT	North or south latitude-current.
6.	CURLAT	Current latitude.
7.	CARLONG	Carrington longitude.
8.	AGE	Age of region in days this transit.

WHITE LIGHT

9.	SPOTCLAS	None observed	0
	Spot Class	Spot class three letter code....................	---
		No data...	9
10.	MAGCLAS	No spots..	0
	Magnetic Class	Alpha...	1
		Beta..	2
		Beta-Gamma......................................	3
		Gamma...	4
		Beta-Delta......................................	5
		Beta-Gamma-Delta................................	6
		Gamma-Delta.....................................	7
		No data...	9
11.	RV	No spots..	0
	Magnetic Field	Red (+ polarity)................................	1
	Strengths	Violet (- polarity).............................	2
	Polarity	No data...	9

12.	MAGSTR	No spots....................................	0
	Magnetic Field	Two digit value.............................	--
	Strengths	(if same use polarity of largest total)	
	(Largest)	No data.....................................	99
13.	MAGGRAD	No spots or unipolar region.................	0.00
	Magnetic	Enter three digit gradient as N.NN..........	-.--
	Gradients	No data.....................................	9.00
	In Gamma/Km		
14.	SSDYNAM	No spots or not applicable..................	0
	Sunspot	Coalescing of spots.........................	1
	Dynamics	Spot rotation...............................	2
		Relative spot motion (opposite polarity spots)..	3
		No data.....................................	9
15.	SSINTER	None occurred...............................	0
	Interaction	Strong spots of opposite polarity converge	
	With Another	(from less than 2 degrees apart)............	1
	Region	No data.....................................	9
16.	STGDEV	No spots....................................	0
	Stage of	Mature group (stable).......................	1
	Development	Decaying....................................	2
		Growing.....................................	3
		Rapid decay (spot or area decrease by > 50%)....	4
		Rapid growth (spot or area increase by > 50%)...	5
		Rapid growth (spot or area increase by > 100%)..	6
		No data.....................................	9

H - ALPHA

17.	LEADTRAI	Structure not definite......................	0
	Leader Emerged	Returning region............................	1
	in Leader or	< 5 deg of NL and out of phase with NL.........	2
	Trailer Polar-	> 5 deg of NL and in leader polarity fields.....	3
	ity Fields	> 5 deg of NL and in trailer polarity fields....	4
	(from Previous	< 5 deg of NL and in-phase with NL.............	5
	Synoptic map)	No data.....................................	9
18.	RETREG	Region not returning........................	0
	Region Number	Region # if returning.......................	----
	if Returning	No data.....................................	9
	Region		

19.	SECTEOW	Sector structure not definite..................	0
	Relationship	Region is > 30 degrees from nearest boundary....	1
	with Nearest	Non-Hale and 10 to 30 deg west of boundary......	2
	Sector Bound-	Non-Hale and 10 to 30 deg east of boundary......	3
	ary (Hale =	Non-Hale and < 10 deg of boundary...............	4
	Region Polarity	Hale and 10 to 30 deg west of boundary..........	5
	Matches the	Hale and 10 to 30 deg east of boundary..........	6
	Boundary)	Hale and < 10 deg of boundary..................	7
		No data.......................................	9
20.	PLAGFIL	Non-compact plage and no filament...............	0
	Plage Compact-	Non-compact plage with filament.................	1
	ness and Embed-	Non-compact plage with active filament..........	2
	ded Filament	Compact plage without embedded filament.........	3
	(Compact = NL	Compact plage with embedded filament............	4
	Corridor > 2	Compact plage with active embedded filament.....	5
	Degrees Wide)	No data.......................................	9
21.	NEUTLOR	Weak structure.................................	0
	Main NL Orien-	North-south (+/- 45 degrees to NS)..............	1
	tation within	East-west......................................	2
	Plage	Hairpin (E-W)..................................	3
		Mostly Circular................................	4
		No data.......................................	9
22.	REVPOL	No reverse polarity............................	0
	Orientation	Reverse Polarity...............................	1
	Within Plage	No data.......................................	9
23.	NEUTLCOM	No kinks or weak structure.....................	0
	Neutral Line	1-3 kinks (very simple region)..................	1
	Complexity	4-6 kinks (simple region)......................	2
		7-12 kinks (intermediate region)................	3
		> 12 kinks (very complex)......................	4
		No data.......................................	9
24.	NEUTLCHG	No definite trend..............................	0
	Neutral Line	Neutral line becoming simple...................	1
	Temporal	Neutral line becoming complex..................	2
	Changes	No data.......................................	9
25.	ASSOCFIL	No associated filament.........................	0
	Associated	Filament unchanged.............................	1
	Filament	Filament growing...............................	2
	(External to	Filament disappeared within past 24 hours.......	3
	Region but	Filament darkens or is active..................	4
	Along Common	No data.......................................	9
	Neutral Line)		

26.	BRTPTS Bright Points	None occurred....................................	0
		Occurred but not along neutral line............	1
		Occurred along the neutral line................	2
		No data...	9
27.	PLAGFLUX Plage Fluctuations	None occurred....................................	0
		Plage fluctuations..............................	1
		No data...	9
28.	ISOPOLE Isolated Pole	None occurred or region is new..................	0
		Isolated pole in region.........................	1
		No data...	9
29.	EFR Emerging Flux	None occurred or region is new..................	0
		New EFR emerges within existing spot group......	1
		New EFR emerges near region (within 5 degrees of existing spot group).....	2
		No data...	9
30.	AFS AFS Present	None present....................................	0
		AFS present.....................................	1
		No data...	9

RADIO

31.	RADIOACT Radio Burst and/or Sweep (Multiple Entries Possible)	None occurred or small events...................	0
		> 250 flux units at 10 cm.......................	1
		>1000 flux units at 10 cm.......................	2
		Type III sweep..................................	3
		Type IV sweep...................................	4
		Type II followed by type IV sweep...............	5
		U Burst...	6
		Major and complex 10 cm burst...................	7
		>1000 flux units at 10 cm and a u burst.........	8
		Type III and type IV sweep......................	10
		> 250 flux units at 10 cm and type III and type IV sweep...........................	11
		No data...	9

HISTORY THIS TRANSIT

32.	FLAREHIS Largest Flare Since Region Appeared Including Today	None occurred or first day observed.............	0
		C class flares have occurred....................	1
		M class flares have occurred....................	2
		X class flares have occurred....................	3
		No data or region appeared on east limb.........	9

33.	FIRSTAPP Region First Appeared	Formed on disk..................................	0
		Came around east limb - first transit...........	1
		Second transit..................................	2
		Third transit...................................	3
		Fourth transit..................................	4
		Fifth transit (and etc).........................	5
		No data...	99
34.	PROTHIS Proton Event	No particle event...............................	0
		Proton 10 event (=10p/cm*cm*sec*ster at >10mev).	1
		Ground level event..............................	2
		No data...	9

REGION FORECASTS

35.	CRFOR Flare Forecast	C probability...................................	
36.	MRFOR Flare Forecast	M probability...................................	
37.	XRFOR Flare Forecast	X probability...................................	
38.	PRFOR Flare Forecast	Proton event probability........................	

EVENTS THAT OCCURRED DURING NEXT 24 HOURS

39.	FLARER Largest Flare for this Region	None occurred or <C0...........................	0
		Class C...	1
		Class M...	2
		Class X...	3
		No data - or region rotated off.................	9
40.	PROTONR Proton Event for this Region	None occurred...................................	0
		Proton event....................................	1
		No data...	9

TOTAL SUN VARIABLES

41.	FLUX 10 Cm Flux	10 cm flux for today............................	

FORECAST FOR SUN

42.	CSFOR Flare Forecast	C probability..................................	
43.	MSFOR Flare Forecast	M probability..................................	
44.	XSFOR Flare Forecast	X probability..................................	
45.	PSFOR Proton Forecast	Proton event probability........................	

EVENTS THAT OCCURRED DURING THIS 24 HOURS

46.	FLARERT	None occurred or <C0...........................	0
	Largest Flare	Class C......................................	1
	for this Region	Class M......................................	2
		Class X......................................	3
		No data or region rotated off..................	9
47.	RECSPOT	No spots observed............................	0
	Recoded Spot	Less than 10%................................	1
	Class	Between 10% and 20%..........................	2
		Between 20% and 30%..........................	3
		Between 30% and 50%..........................	4
		Between 50% and 60%..........................	5
		Between 60% and 100%.........................	6
		Between 100% and 200%........................	7
		Between 200% and 300%........................	8
		Spotclas didn't occur in last eight years.......	98
		No data......................................	99
48.	PROTONT	None occurred................................	0
	Proton Event	Proton event.................................	1
	for this	No data......................................	9
	Region		

APPENDIX B

Probability Forecast Tables

The purpose of this appendix is to present classification tables for SESC, LR and DA probability forecasts, based on the analysis presented in section 5 of this report. Where applicable, FLARER is crosstabulated with appropriate probability estimates using the SPSS package of computer programs. Variables crosstabulated with FLARER are identified by a one letter source code followed by one to three letters indicating the event which is forecast. For example, SCMX denotes the SESC probability of a C, M or X flare occurring; i.e., SCMX is equivalent to CRFOR in the data base. Thus LMX is the LR forecast of an M or X event, analogous to MRFOR in the data base.

Tables reported in B.1 - B.3 are for Model II. Results for Model IV are given in B.4.

Contents

B.0 SESC Overall Classification

B.1 No Flare in Past 24 Hours

- B.1.1 Probability of C, M, or X
- B.1.2 Probability of M or X
- B.1.3 Probability of X

B.2 C Flare in Past 24 Hours

- B.2.1 Probability of C, M, or X
- B.2.2 Probability of M or X
- B.2.3 Probability of X

B.3 M Flare in Past 24 Hours

- B.3.1 Probability of C, M, or X
- B.3.2 Probability of M or X
- B.3.3 Probability of X

B.4 X Flare in Past 24 Hours

- B.4.1 Probability of C, M, or X
- B.4.2 Probability of M or X
- B.4.3 Probability of X

B.0 SESC Overall Classification

CRFOR

FLARER (COUNT / ROW PCT)	0 - 10 (1)	10 - 20 (2)	20 - 30 (3)	30 - 40 (4)	40 - 50 (5)	50 - 60 (6)	60 - 70 (7)	70 - 80 (8)	80 - 90 (9)	90 -100 (10)	ROW TOTAL
0 NO FLARE NXT DAY	2171	426	270	194	202	101	80	115	132	27	3718
	58.4	11.5	7.3	5.2	5.4	2.7	2.2	3.1	3.6	.7	82.9
1 C CLASS NEXT DAY	110	46	49	38	43	40	33	54	127	46	586
	18.8	7.8	8.4	6.5	7.3	6.8	5.6	9.2	21.7	7.8	13.1
2 M CLASS NEXT DAY	16	2	8	4	8	10	8	19	51	35	161
	9.9	1.2	5.0	2.5	5.0	6.2	5.0	11.8	31.7	21.7	3.6
3 X CLASS NEXT DAY	0	0	1	0	0	2	0	1	13	5	22
	0	0	4.5	0	0	9.1	0	4.5	59.1	22.7	.5
COLUMN TOTAL	2297	474	328	236	253	153	121	189	323	113	4487
	51.2	10.6	7.3	5.3	5.6	3.4	2.7	4.2	7.2	2.5	100.0

KENDALL'S TAU C = .22981. SIGNIFICANCE = 0
SOMERS'S D (ASYMMETRIC) = .24413 WITH FLARER DEPENDENT. = .50420 WITH CRFOR DEPENDENT.
SOMERS'S D (SYMMETRIC) = .34436

MRFOR

FLARER (COUNT / ROW PCT)	0 - 10 (1)	10 - 20 (2)	20 - 30 (3)	30 - 40 (4)	40 - 50 (5)	50 - 60 (6)	60 - 70 (7)	70 - 80 (8)	80 - 90 (9)	90 -100 (10)	ROW TOTAL
0 NO FLARE NXT DAY	3293	187	97	48	44	26	11	7	4	1	3718
	88.6	5.0	2.6	1.3	1.2	.7	.3	.2	.1	.0	82.9
1 C CLASS NEXT DAY	314	81	56	33	38	14	20	14	10	6	586
	53.6	13.8	9.6	5.6	6.5	2.4	3.4	2.4	1.7	1.0	13.1
2 M CLASS NEXT DAY	44	18	22	14	13	8	13	12	11	6	161
	27.3	11.2	13.7	8.7	8.1	5.0	8.1	7.5	6.8	3.7	3.6
3 X CLASS NEXT DAY	1	1	1	2	5	3	3	1	3	2	22
	4.5	4.5	4.5	9.1	22.7	13.6	13.6	4.5	13.6	9.1	.5
COLUMN TOTAL	3652	287	176	97	100	51	47	34	28	15	4487
	81.4	6.4	3.9	2.2	2.2	1.1	1.0	.8	.6	.3	100.0

KENDALL'S TAU C = .17165. SIGNIFICANCE = 0
SOMERS'S D (ASYMMETRIC) = .38939 WITH FLARER DEPENDENT. = .43636 WITH MRFOR DEPENDENT.
SOMERS'S D (SYMMETRIC) = .41194

XRFOR

FLARER (COUNT / ROW PCT)	0 - 10 (1)	10 - 20 (2)	20 - 30 (3)	30 - 40 (4)	40 - 50 (5)	50 - 60 (6)	60 - 70 (7)	ROW TOTAL
0 NO FLARE NXT DAY	3675	34	5	3	1	0	0	3718
	98.8	.9	.1	.1	.0	0	0	82.9
1 C CLASS NEXT DAY	528	31	12	5	8	0	2	586
	90.1	5.3	2.0	.9	1.4	0	.3	13.1
2 M CLASS NEXT DAY	114	15	18	4	8	2	0	161
	70.8	9.3	11.2	2.5	5.0	1.2	0	3.6
3 X CLASS NEXT DAY	11	2	7	0	2	0	0	22
	50.0	9.1	31.8	0	9.1	0	0	.5
COLUMN TOTAL	4328	82	42	12	19	2	2	4487
	96.5	1.8	.9	.3	.4	.0	.0	100.0

KENDALL'S TAU C = .05620. SIGNIFICANCE = 0
SOMERS'S D (ASYMMETRIC) = .61020 WITH FLARER DEPENDENT. = .14300 WITH XRFOR DEPENDENT.
SOMERS'S D (SYMMETRIC) = .23101

B.1 No Flare in Past 24 Hours

B.1.1 Probability of C, M, or X

SCMX

FLARER	COUNT / ROW PCT	0 - 10 (1)	10 - 20 (2)	20 - 30 (3)	30 - 40 (4)	40 - 50 (5)	50 - 60 (6)	60 - 70 (7)	70 - 80 (8)	80 - 90 (9)	90 -100 (10)	ROW TOTAL
NO FLARE	0	997	222	147	105	98	50	33	58	47	4	1761
		56.6	12.6	8.3	6.0	5.6	2.8	1.9	3.3	2.7	.2	88.3
C FLARE	1	48	23	26	18	17	19	10	17	24	4	206
		23.3	11.2	12.6	8.7	8.3	9.2	4.9	8.3	11.7	1.9	10.3
M FLARE	2	6	1	2	0	4	4	1	3	5	0	26
		23.1	3.8	7.7	0	15.4	15.4	3.8	11.5	19.2	0	1.3
X FLARE	3	0	0	0	0	0	0	0	0	1	0	1
		0	0	0	0	0	0	0	0	100.0	0	.1
	COLUMN TOTAL	1051	246	175	123	119	73	44	78	77	8	1994
		52.7	12.3	8.8	6.2	6.0	3.7	2.2	3.9	3.9	.4	100.0

LCMX

FLARER	COUNT / ROW PCT	0 - 10 (1)	10 - 20 (2)	20 - 30 (3)	30 - 40 (4)	40 - 50 (5)	50 - 60 (6)	60 - 70 (7)	80 - 90 (9)	ROW TOTAL
NO FLARE	0	1120	462	98	46	17	15	2	1	1761
		63.6	26.2	5.6	2.6	1.0	.9	.1	.1	88.3
C FLARE	1	59	69	31	23	16	5	3	0	206
		28.6	33.5	15.0	11.2	7.8	2.4	1.5	0	10.3
M FLARE	2	7	6	4	5	2	1	1	0	26
		26.9	23.1	15.4	19.2	7.7	3.8	3.8	0	1.3
X FLARE	3	0	0	1	0	0	0	0	0	1
		0	0	100.0	0	0	0	0	0	.1
	COLUMN TOTAL	1186	537	134	74	35	21	6	1	1994
		59.5	26.9	6.7	3.7	1.8	1.1	.3	.1	100.0

OCMX

FLARER	COUNT / ROW PCT	0 - 10 (1)	10 - 20 (2)	20 - 30 (3)	30 - 40 (4)	40 - 50 (5)	50 - 60 (6)	60 - 70 (7)	70 - 80 (8)	80 - 90 (9)	90 -100 (10)	ROW TOTAL
NO FLARE	0	1218	334	95	44	27	20	13	6	2	2	1761
		69.2	19.0	5.4	2.5	1.5	1.1	.7	.3	.1	.1	88.3
C FLARE	1	73	51	22	20	16	10	10	4	0	0	206
		35.4	24.8	10.7	9.7	7.8	4.9	4.9	1.9	0	0	10.3
M FLARE	2	8	4	3	4	1	4	0	1	1	0	26
		30.8	15.4	11.5	15.4	3.8	15.4	0	3.8	3.8	0	1.3
X FLARE	3	0	0	1	0	0	0	0	0	0	0	1
		0	0	100.0	0	0	0	0	0	0	0	.1
	COLUMN TOTAL	1299	389	121	68	44	34	23	11	3	2	1994
		65.1	19.5	6.1	3.4	2.2	1.7	1.2	.6	.2	.1	100.0

B.1.2 Probability of M or X

SMX

FLARER (COUNT / ROW PCT)		0 - 10 (1)	10 - 20 (2)	20 - 30 (3)	30 - 40 (4)	40 - 50 (5)	50 - 60 (6)	60 - 70 (7)	70 - 80 (8)	ROW TOTAL
NO FLARE	0	1595	83	35	22	13	6	6	1	1761
		90.6	4.7	2.0	1.2	.7	.3	.3	.1	88.3
C FLARE	1	149	23	10	9	9	1	3	2	206
		72.3	11.2	4.9	4.4	4.4	.5	1.5	1.0	10.3
M FLARE	2	18	3	2	0	0	2	1	0	26
		69.2	11.5	7.7	0	0	7.7	3.8	0	1.3
X FLARE	3	0	0	0	0	0	0	0	1	1
		0	0	0	0	0	0	0	100.0	.1
COLUMN TOTAL		1762	109	47	31	22	9	10	4	1994
		88.4	5.5	2.4	1.6	1.1	.5	.5	.2	100.0

LMX

FLARER (COUNT / ROW PCT)		0 - 10 (1)	10 - 20 (2)	20 - 30 (3)	30 - 40 (4)	ROW TOTAL
NO FLARE	0	1749	10	1	1	1761
		99.3	.6	.1	.1	88.3
C FLARE	1	200	6	0	0	206
		97.1	2.9	0	0	10.3
M FLARE	2	23	3	0	0	26
		88.5	11.5	0	0	1.3
X FLARE	3	1	0	0	0	1
		100.0	0	0	0	.1
COLUMN TOTAL		1973	19	1	1	1994
		98.9	1.0	.1	.1	100.0

DMX

FLARER (COUNT / ROW PCT)		0 - 10 (1)	10 - 20 (2)	20 - 30 (3)	30 - 40 (4)	40 - 50 (5)	80 - 90 (9)	90 -100 (10)	ROW TOTAL
NO FLARE	0	1697	41	8	7	4	2	2	1761
		96.4	2.3	.5	.4	.2	.1	.1	88.3
C FLARE	1	178	11	8	6	3	0	0	206
		86.4	5.3	3.9	2.9	1.5	0	0	10.3
M FLARE	2	16	5	1	2	2	0	0	26
		61.5	19.2	3.8	7.7	7.7	0	0	1.3
X FLARE	3	1	0	0	0	0	0	0	1
		100.0	0	0	0	0	0	0	.1
COLUMN TOTAL		1892	57	17	15	9	2	2	1994
		94.9	2.9	.9	.8	.5	.1	.1	100.0

B.1.3 Probability of X

FLARER	COUNT / ROW PCT	SX: 0 - 10 (1)	10 - 20 (2)	30 - 40 (4)	ROW TOTAL
NO FLARE	0	1750	10	1	1761
		99.4	.6	.1	88.3
C FLARE	1	201	5	0	206
		97.6	2.4	0	10.3
M FLARE	2	26	0	0	26
		100.0	0	0	1.3
X FLARE	3	0	1	0	1
		0	100.0	0	.1
	COLUMN TOTAL	1977	16	1	1994
		99.1	.8	.1	100.0

B.2 C Flare in Past 24 Hours

B.2.1 Probability of C, M, or X

SCMX

FLARER (COUNT / ROW PCT)	0 - 10 (1)	10 - 20 (2)	20 - 30 (3)	30 - 40 (4)	40 - 50 (5)	50 - 60 (6)	60 - 70 (7)	70 - 80 (8)	80 - 90 (9)	90 -100 (10)	ROW TOTAL
0 NO FLARE	47	19	24	26	30	26	22	26	38	12	270
	17.4	7.0	8.9	9.6	11.1	9.6	8.1	9.6	14.1	4.4	52.8
1 C FLARE	13	3	10	11	12	13	17	22	68	17	186
	7.0	1.6	5.4	5.9	6.5	7.0	9.1	11.8	36.6	9.1	36.4
2 M FLARE	1	1	0	2	2	3	2	9	19	8	47
	2.1	2.1	0	4.3	4.3	6.4	4.3	19.1	40.4	17.0	9.2
3 X FLARE	0	0	1	0	0	1	0	0	6	0	8
	0	0	12.5	0	0	12.5	0	0	75.0	0	1.6
COLUMN TOTAL	61	23	35	39	44	43	41	57	131	37	511
	11.9	4.5	6.8	7.6	8.6	8.4	8.0	11.2	25.6	7.2	100.0

LCMX

FLARER (COUNT / ROW PCT)	10 - 20 (2)	20 - 30 (3)	30 - 40 (4)	40 - 50 (5)	50 - 60 (6)	60 - 70 (7)	70 - 80 (8)	80 - 90 (9)	90 -100 (10)	ROW TOTAL
0 NO FLARE	28	69	64	32	35	28	10	4	0	270
	10.4	25.6	23.7	11.9	13.0	10.4	3.7	1.5	0	52.8
1 C FLARE	8	14	25	28	32	29	31	16	3	186
	4.3	7.5	13.4	15.1	17.2	15.6	16.7	8.6	1.6	36.4
2 M FLARE	0	4	2	8	10	5	6	9	3	47
	0	8.5	4.3	17.0	21.3	10.6	12.8	19.1	6.4	9.2
3 X FLARE	0	0	2	2	1	1	1	1	0	8
	0	0	25.0	25.0	12.5	12.5	12.5	12.5	0	1.6
COLUMN TOTAL	36	87	93	70	78	63	48	30	6	511
	7.0	17.0	18.2	13.7	15.3	12.3	9.4	5.9	1.2	100.0

DCMX

FLARER (COUNT / ROW PCT)	0 - 10 (1)	10 - 20 (2)	20 - 30 (3)	30 - 40 (4)	40 - 50 (5)	50 - 60 (6)	60 - 70 (7)	70 - 80 (8)	80 - 90 (9)	90 -100 (10)	ROW TOTAL
0 NO FLARE	6	27	76	54	32	33	25	13	4	0	270
	2.2	10.0	28.1	20.0	11.9	12.2	9.3	4.8	1.5	0	52.8
1 C FLARE	0	8	17	22	30	29	30	31	16	3	186
	0	4.3	9.1	11.8	16.1	15.6	16.1	16.7	8.6	1.6	36.4
2 M FLARE	0	0	4	3	7	10	5	5	10	3	47
	0	0	8.5	6.4	14.9	21.3	10.6	10.6	21.3	6.4	9.2
3 X FLARE	0	0	0	2	2	1	0	2	1	0	8
	0	0	0	25.0	25.0	12.5	0	25.0	12.5	0	1.6
COLUMN TOTAL	6	35	97	81	71	73	60	51	31	6	511
	1.2	6.8	19.0	15.9	13.9	14.3	11.7	10.0	6.1	1.2	100.0

B.2.2 Probability of M or X

SMX

FLARER	COUNT / ROW PCT	0 - 10 (1)	10 - 20 (2)	20 - 30 (3)	30 - 40 (4)	40 - 50 (5)	50 - 60 (6)	60 - 70 (7)	70 - 80 (8)	80 - 90 (9)	90 - 100 (10)	ROW TOTAL
NO FLARE	0	164	39	26	11	14	10	3	3	0	0	270
		60.7	14.4	9.6	4.1	5.2	3.7	1.1	1.1	0	0	52.8
C FLARE	1	61	34	34	16	14	5	11	7	2	2	186
		32.8	18.3	18.3	8.6	7.5	2.7	5.9	3.8	1.1	1.1	36.4
M FLARE	2	8	7	13	4	6	2	1	4	1	1	47
		17.0	14.9	27.7	8.5	12.8	4.3	2.1	8.5	2.1	2.1	9.2
X FLARE	3	1	1	1	0	2	2	1	0	0	0	8
		12.5	12.5	12.5	0	25.0	25.0	12.5	0	0	0	1.6
COLUMN TOTAL		234	81	74	31	36	19	16	14	3	3	511
		45.8	15.9	14.5	6.1	7.0	3.7	3.1	2.7	.6	.6	100.0

LMX

FLARER	COUNT / ROW PCT	0 - 10 (1)	10 - 20 (2)	20 - 30 (3)	30 - 40 (4)	40 - 50 (5)	50 - 60 (6)	ROW TOTAL
NO FLARE	0	204	58	5	3	0	0	270
		75.6	21.5	1.9	1.1	0	0	52.8
C FLARE	1	87	72	11	13	2	1	186
		46.8	38.7	5.9	7.0	1.1	.5	36.4
M FLARE	2	14	19	2	8	4	0	47
		29.8	40.4	4.3	17.0	8.5	0	9.2
X FLARE	3	4	2	0	2	0	0	8
		50.0	25.0	0	25.0	0	0	1.6
COLUMN TOTAL		309	151	18	26	6	1	511
		60.5	29.5	3.5	5.1	1.2	.2	100.0

DMX

FLARER	COUNT / ROW PCT	0 - 10 (1)	10 - 20 (2)	20 - 30 (3)	30 - 40 (4)	40 - 50 (5)	50 - 60 (6)	70 - 80 (8)	ROW TOTAL
NO FLARE	0	211	51	3	3	2	0	0	270
		78.1	18.9	1.1	1.1	.7	0	0	52.8
C FLARE	1	88	71	2	15	7	2	1	186
		47.3	38.2	1.1	8.1	3.8	1.1	.5	36.4
M FLARE	2	15	16	4	3	7	2	0	47
		31.9	34.0	8.5	6.4	14.9	4.3	0	9.2
X FLARE	3	5	1	0	1	1	0	0	8
		62.5	12.5	0	12.5	12.5	0	0	1.6
COLUMN TOTAL		319	139	9	22	17	4	1	511
		62.4	27.2	1.8	4.3	3.3	.8	.2	100.0

B.2.3 Probability of X

FLARER	COUNT / ROW PCT	SX 0 - 10 (1)	10 - 20 (2)	20 - 30 (3)	30 - 40 (4)	40 - 50 (5)	60 - 70 (7)	ROW TOTAL
NO FLARE	0	258	10	2	0	0	0	270
		95.6	3.7	.7	0	0	0	52.8
C FLARE	1	162	14	4	3	2	1	186
		87.1	7.5	2.2	1.6	1.1	.5	36.4
M FLARE	2	37	3	4	1	2	0	47
		78.7	6.4	8.5	2.1	4.3	0	9.2
X FLARE	3	6	0	2	0	0	0	8
		75.0	0	25.0	0	0	0	1.6
COLUMN TOTAL		463	27	12	4	4	1	511
		90.6	5.3	2.3	.8	.8	.2	100.0

B.3 M Flare in Past 24 Hours

B.3.1 Probability of C, M, or X

SCMX

FLARER	COUNT / ROW PCT	0 - 10 (1)	20 - 30 (3)	30 - 40 (4)	40 - 50 (5)	50 - 60 (6)	60 - 70 (7)	70 - 80 (8)	80 - 90 (9)	90 -100 (10)	ROW TOTAL
NO FLARE (0)	COUNT	5	1	3	4	2	3	7	17	5	47
	ROW PCT	10.6	2.1	6.4	8.5	4.3	6.4	14.9	36.2	10.6	35.3
C FLARE (1)	COUNT	0	1	0	1	2	2	6	18	14	44
	ROW PCT	0	2.3	0	2.3	4.5	4.5	13.6	40.9	31.8	33.1
M FLARE (2)	COUNT	0	0	0	0	2	1	2	15	18	38
	ROW PCT	0	0	0	0	5.3	2.6	5.3	39.5	47.4	28.6
X FLARE (3)	COUNT	0	0	0	0	0	0	1	1	2	4
	ROW PCT	0	0	0	0	0	0	25.0	25.0	50.0	3.0
COLUMN	TOTAL	5	2	3	5	6	6	16	51	39	133
		3.8	1.5	2.3	3.8	4.5	4.5	12.0	38.3	29.3	100.0

LCMX

FLARER	COUNT / ROW PCT	0 - 10 (1)	10 - 20 (2)	20 - 30 (3)	30 - 40 (4)	40 - 50 (5)	50 - 60 (6)	60 - 70 (7)	70 - 80 (8)	80 - 90 (9)	90 -100 (10)	ROW TOTAL
NO FLARE (0)	COUNT	4	4	8	7	7	5	3	5	2	2	47
	ROW PCT	8.5	8.5	17.0	14.9	14.9	10.6	6.4	10.6	4.3	4.3	35.3
C FLARE (1)	COUNT	0	2	1	1	3	4	7	4	6	16	44
	ROW PCT	0	4.5	2.3	2.3	6.8	9.1	15.9	9.1	13.6	36.4	33.1
M FLARE (2)	COUNT	0	1	1	0	2	3	3	7	6	15	38
	ROW PCT	0	2.6	2.6	0	5.3	7.9	7.9	18.4	15.8	39.5	28.6
X FLARE (3)	COUNT	0	0	0	0	0	0	0	0	0	4	4
	ROW PCT	0	0	0	0	0	0	0	0	0	100.0	3.0
COLUMN	TOTAL	4	7	10	8	12	12	13	16	14	37	133
		3.0	5.3	7.5	6.0	9.0	9.0	9.8	12.0	10.5	27.8	100.0

DCMX

FLARER	COUNT / ROW PCT	0 - 10 (1)	10 - 20 (2)	20 - 30 (3)	30 - 40 (4)	40 - 50 (5)	50 - 60 (6)	60 - 70 (7)	70 - 80 (8)	80 - 90 (9)	90 -100 (10)	ROW TOTAL
NO FLARE (0)	COUNT	5	6	6	7	5	6	4	2	4	2	47
	ROW PCT	10.6	12.8	12.8	14.9	10.6	12.8	8.5	4.3	8.5	4.3	35.3
C FLARE (1)	COUNT	0	2	2	0	3	5	7	2	8	15	44
	ROW PCT	0	4.5	4.5	0	6.8	11.4	15.9	4.5	18.2	34.1	33.1
M FLARE (2)	COUNT	0	1	1	1	1	5	2	6	6	15	38
	ROW PCT	0	2.6	2.6	2.6	2.6	13.2	5.3	15.8	15.8	39.5	28.6
X FLARE (3)	COUNT	0	0	0	0	0	0	0	0	0	4	4
	ROW PCT	0	0	0	0	0	0	0	0	0	100.0	3.0
COLUMN	TOTAL	5	9	9	8	9	16	13	10	18	36	133
		3.8	6.8	6.8	6.0	6.8	12.0	9.8	7.5	13.5	27.1	100.0

B.3.2 Probability of M or X

SMX

FLARER	COUNT / ROW PCT	0 - 10 (1)	10 - 20 (2)	20 - 30 (3)	30 - 40 (4)	40 - 50 (5)	50 - 60 (6)	60 - 70 (7)	70 - 80 (8)	80 - 90 (9)	90 - 100 (10)	ROW TOTAL
NO FLARE	0	13	9	6	5	6	5	1	1	1	0	47
		27.7	19.1	12.8	10.6	12.8	10.6	2.1	2.1	2.1	0	35.3
C FLARE	1	3	3	4	5	12	4	2	3	4	4	44
		6.8	6.8	9.1	11.4	27.3	9.1	4.5	6.8	9.1	9.1	33.1
M FLARE	2	0	1	4	4	7	1	8	5	5	3	38
		0	2.6	10.5	10.5	18.4	2.6	21.1	13.2	13.2	7.9	28.6
X FLARE	3	0	0	0	0	3	0	1	0	0	0	4
		0	0	0	0	75.0	0	25.0	0	0	0	3.0
COLUMN TOTAL		16	13	14	14	28	10	12	9	10	7	133
		12.0	9.8	10.5	10.5	21.1	7.5	9.0	6.8	7.5	5.3	100.0

LMX

FLARER	COUNT / ROW PCT	0 - 10 (1)	10 - 20 (2)	20 - 30 (3)	30 - 40 (4)	40 - 50 (5)	50 - 60 (6)	60 - 70 (7)	70 - 80 (8)	ROW TOTAL
NO FLARE	0	15	15	10	1	3	2	1	0	47
		31.9	31.9	21.3	2.1	6.4	4.3	2.1	0	35.3
C FLARE	1	7	8	5	9	6	6	2	1	44
		15.9	18.2	11.4	20.5	13.6	13.6	4.5	2.3	33.1
M FLARE	2	3	3	5	6	6	4	5	6	38
		7.9	7.9	13.2	15.8	15.8	10.5	13.2	15.8	28.6
X FLARE	3	0	0	0	0	1	2	0	1	4
		0	0	0	0	25.0	50.0	0	25.0	3.0
COLUMN TOTAL		25	26	20	16	16	14	8	8	133
		18.8	19.5	15.0	12.0	12.0	10.5	6.0	6.0	100.0

DMX

FLARER	COUNT / ROW PCT	0 - 10 (1)	10 - 20 (2)	20 - 30 (3)	30 - 40 (4)	40 - 50 (5)	50 - 60 (6)	60 - 70 (7)	70 - 80 (8)	ROW TOTAL
NO FLARE	0	13	18	10	1	3	1	1	0	47
		27.7	38.3	21.3	2.1	6.4	2.1	2.1	0	35.3
C FLARE	1	7	8	6	7	7	4	4	1	44
		15.9	18.2	13.6	15.9	15.9	9.1	9.1	2.3	33.1
M FLARE	2	2	4	5	6	6	4	4	7	38
		5.3	10.5	13.2	15.8	15.8	10.5	10.5	18.4	28.6
X FLARE	3	0	0	0	0	1	1	1	1	4
		0	0	0	0	25.0	25.0	25.0	25.0	3.0
COLUMN TOTAL		22	30	21	14	17	10	10	9	133
		16.5	22.6	15.8	10.5	12.8	7.5	7.5	6.8	100.0

B.3.3 Probability of X

FLARER	COUNT / ROW PCT	SX 0 - 10 (1)	10 - 20 (2)	20 - 30 (3)	30 - 40 (4)	40 - 50 (5)	50 - 60 (6)	60 - 70 (7)	ROW TOTAL
NO FLARE	0	39	7	1	0	0	0	0	47
		83.0	14.9	2.1	0	0	0	0	35.3
C FLARE	1	28	3	5	1	6	0	1	44
		63.6	6.8	11.4	2.3	13.6	0	2.3	33.1
M FLARE	2	17	6	9	3	2	1	0	38
		44.7	15.8	23.7	7.9	5.3	2.6	0	28.6
X FLARE	3	2	1	1	0	0	0	0	4
		50.0	25.0	25.0	0	0	0	0	3.0
COLUMN TOTAL		86	17	16	4	8	1	1	133
		64.7	12.8	12.0	3.0	6.0	.8	.8	100.0

B.4 X Flare in Past 24 Hours

B.4.1 Probability of C, M, or X

SCMX

FLARER	COUNT / ROW PCT	30 - 40 (4)	50 - 60 (6)	80 - 90 (9)	90 -100 (10)	ROW TOTAL
NO FLARE	0	0 0	1 33.3	1 33.3	1 33.3	3 14.3
C FLARE	1	0 0	0 0	3 60.0	2 40.0	5 23.8
M FLARE	2	1 14.3	0 0	2 28.6	4 57.1	7 33.3
X FLARE	3	0 0	0 0	3 50.0	3 50.0	6 28.6
	COLUMN TOTAL	1 4.8	1 4.8	9 42.9	10 47.6	21 100.0

LCMX

FLARER	COUNT / ROW PCT	0 - 10 (1)	60 - 70 (7)	90 -100 (10)	ROW TOTAL
NO FLARE	0	2 66.7	1 33.3	0 0	3 14.3
C FLARE	1	0 0	1 20.0	4 80.0	5 23.8
M FLARE	2	0 0	1 14.3	6 85.7	7 33.3
X FLARE	3	0 0	0 0	6 100.0	6 28.6
	COLUMN TOTAL	2 9.5	3 14.3	16 76.2	21 100.0

DCMX

FLARER	COUNT / ROW PCT	0 - 10 (1)	40 - 50 (5)	70 - 80 (8)	90 -100 (10)	ROW TOTAL
NO FLARE	0	2 66.7	0 0	1 33.3	0 0	3 14.3
C FLARE	1	0 0	1 20.0	1 20.0	3 60.0	5 23.8
M FLARE	2	0 0	0 0	1 14.3	6 85.7	7 33.3
X FLARE	3	0 0	0 0	0 0	6 100.0	6 28.6
	COLUMN TOTAL	2 9.5	1 4.8	3 14.3	15 71.4	21 100.0

B.4.2 Probability of M or X

FLARER	COUNT / ROW PCT	SMX 30 - 40 (4)	40 - 50 (5)	50 - 60 (6)	60 - 70 (7)	70 - 80 (8)	80 - 90 (9)	90 -100 (10)	ROW TOTAL
NO FLARE	0	1	1	0	0	0	0	1	3
		33.3	33.3	0	0	0	0	33.3	14.3
C FLARE	1	0	0	0	3	0	2	0	5
		0	0	0	60.0	0	40.0	0	23.8
M FLARE	2	0	0	1	2	1	1	2	7
		0	0	14.3	28.6	14.3	14.3	28.6	33.3
X FLARE	3	0	0	0	1	0	3	2	6
		0	0	0	16.7	0	50.0	33.3	28.6
COLUMN TOTAL		1	1	1	6	1	6	5	21
		4.8	4.8	4.8	28.6	4.8	28.6	23.8	100.0

FLARER	COUNT / ROW PCT	LMX 10 - 20 (2)	30 - 40 (4)	40 - 50 (5)	60 - 70 (7)	70 - 80 (8)	80 - 90 (9)	90 -100 (10)	ROW TOTAL
NO FLARE	0	3	0	0	0	0	0	0	3
		100.0	0	0	0	0	0	0	14.3
C FLARE	1	1	1	1	1	0	0	1	5
		20.0	20.0	20.0	20.0	0	0	20.0	23.8
M FLARE	2	1	0	0	1	2	2	1	7
		14.3	0	0	14.3	28.6	28.6	14.3	33.3
X FLARE	3	0	0	0	1	1	2	2	6
		0	0	0	16.7	16.7	33.3	33.3	28.6
COLUMN TOTAL		5	1	1	3	3	4	4	21
		23.8	4.8	4.8	14.3	14.3	19.0	19.0	100.0

FLARER	COUNT / ROW PCT	DMX 0 - 10 (1)	10 - 20 (2)	20 - 30 (3)	30 - 40 (4)	60 - 70 (7)	80 - 90 (9)	90 -100 (10)	ROW TOTAL
NO FLARE	0	1	2	0	0	0	0	0	3
		33.3	66.7	0	0	0	0	0	14.3
C FLARE	1	1	0	1	1	1	0	1	5
		20.0	0	20.0	20.0	20.0	0	20.0	23.8
M FLARE	2	1	0	0	0	1	2	3	7
		14.3	0	0	0	14.3	28.6	42.9	33.3
X FLARE	3	0	0	0	0	1	1	4	6
		0	0	0	0	16.7	16.7	66.7	28.6
COLUMN TOTAL		3	2	1	1	3	3	8	21
		14.3	9.5	4.8	4.8	14.3	14.3	38.1	100.0

B.4.3 Probability of X

FLARER	SX	0 - 10	10 - 20	20 - 30	40 - 50	50 - 60	ROW TOTAL
COUNT / ROW PCT		1	2	3	5	6	
NO FLARE	0	0	1	2	0	0	3
		0	33.3	66.7	0	0	14.3
C FLARE	1	1	2	2	0	0	5
		20.0	40.0	40.0	0	0	23.8
M FLARE	2	1	2	1	2	1	7
		14.3	28.6	14.3	28.6	14.3	33.3
X FLARE	3	1	0	3	2	0	6
		16.7	0	50.0	33.3	0	28.6
COLUMN TOTAL		3	5	8	4	1	21
		14.3	23.8	38.1	19.0	4.8	100.0

FLARER	LX	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	ROW TOTAL
COUNT / ROW PCT		1	2	3	4	5	6	
NO FLARE	0	3	0	0	0	0	0	3
		100.0	0	0	0	0	0	14.3
C FLARE	1	3	1	0	0	0	1	5
		60.0	20.0	0	0	0	20.0	23.8
M FLARE	2	2	0	0	2	2	1	7
		28.6	0	0	28.6	28.6	14.3	33.3
X FLARE	3	0	0	1	1	2	2	6
		0	0	16.7	16.7	33.3	33.3	28.6
COLUMN TOTAL		8	1	1	3	4	4	21
		38.1	4.8	4.8	14.3	19.0	19.0	100.0

FLARER	DX	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	ROW TOTAL
COUNT / ROW PCT		1	2	3	4	5	6	
NO FLARE	0	3	0	0	0	0	0	3
		100.0	0	0	0	0	0	14.3
C FLARE	1	3	1	0	0	0	1	5
		60.0	20.0	0	0	0	20.0	23.8
M FLARE	2	2	0	0	2	2	1	7
		28.6	0	0	28.6	28.6	14.3	33.3
X FLARE	3	0	0	1	1	2	2	6
		0	0	16.7	16.7	33.3	33.3	28.6
COLUMN TOTAL		8	1	1	3	4	4	21
		38.1	4.8	4.8	14.3	19.0	19.0	100.0

APPENDIX C

Crosstabulations

In this appendix crosstabulations of most solar flare variables with FLARER are presented. For these tables the data are not partitioned on the past flare event, but conditional tables can be easily obtained using the SPSS programs. The reader is referred to the SPSS user's manual for a discussion of the statistics listed below tables. Variables have been grouped according to the categories noted in Appendix A.

Contents

C.1 Location Variables

C.2 White Light Variables

C.3 H-Alpha Variables

C.4 Historical Variables

C.5 Total Sun Variables

C.6 Events During This 24 Hours

C.1 Location Variables

DATE

FLARER	COUNT / ROW PCT / COL PCT / TOT PCT	7JAN-MAR 1	7APR-JUN 2	7JUL-SEP 3	7OCT-DEC 4	8JAN-MAR 5	8APR-JUN 6	8JUL-SEP 7	8OCT-DEC 8	9JAN 9	ROW TOTAL
0 NO FLARE NXT DAY	COUNT	127	134	236	295	501	592	596	864	373	3718
	ROW PCT	3.4	3.6	6.3	7.9	13.5	15.9	16.0	23.2	10.0	82.9
	COL PCT	90.1	80.7	87.7	82.6	83.2	76.3	82.2	85.7	84.2	
	TOT PCT	2.8	3.0	5.3	6.6	11.2	13.2	13.3	19.3	8.3	
1 C CLASS NEXT DAY	COUNT	12	23	24	44	80	136	103	112	52	586
	ROW PCT	2.0	3.9	4.1	7.5	13.7	23.2	17.6	19.1	8.9	13.1
	COL PCT	8.5	13.9	8.9	12.3	13.3	17.5	14.2	11.1	11.7	
	TOT PCT	.3	.5	.5	1.0	1.8	3.0	2.3	2.5	1.2	
2 M CLASS NEXT DAY	COUNT	2	9	7	16	20	41	22	27	17	161
	ROW PCT	1.2	5.6	4.3	9.9	12.4	25.5	13.7	16.8	10.6	3.6
	COL PCT	1.4	5.4	2.6	4.5	3.3	5.3	3.0	2.7	3.8	
	TOT PCT	.0	.2	.2	.4	.4	.9	.5	.6	.4	
3 X CLASS NEXT DAY	COUNT	0	0	2	2	1	7	4	5	1	22
	ROW PCT	0	0	9.1	9.1	4.5	31.8	18.2	22.7	4.5	.5
	COL PCT	0	0	.7	.6	.2	.9	.6	.5	.2	
	TOT PCT	0	0	.0	.0	.0	.2	.1	.1	.0	
COLUMN TOTAL		141	166	269	357	602	776	725	1008	443	4487
		3.1	3.7	6.0	8.0	13.4	17.3	16.2	22.5	9.9	100.0

RAW CHI SQUARE = 49.88377 WITH 24 DEGREES OF FREEDOM. SIGNIFICANCE = .0015
KENDALL#S TAU C = -.00673. SIGNIFICANCE = .2165
GAMMA = -.02005
SOMERS#S D (ASYMMETRIC) = -.00591 WITH FLARER DEPENDENT. = -.01711 WITH DATE DEPENDENT.
SOMERS#S D (SYMMETRIC) = -.00879

APPLONG

FLARER	COUNT / ROW PCT / COL PCT / TOT PCT	0 - 10 1	11-120 2	121-150 3	151-180 4	181-210 5	OVER 210 6	ROW TOTAL
0 NO FLARE NXT DAY	COUNT	1995	308	527	413	145	330	3718
	ROW PCT	53.7	8.3	14.2	11.1	3.9	8.9	82.9
	COL PCT	77.7	86.8	88.3	90.2	92.9	93.8	
	TOT PCT	44.5	6.9	11.7	9.2	3.2	7.4	
1 C CLASS NEXT DAY	COUNT	412	38	67	40	8	21	586
	ROW PCT	70.3	6.5	11.4	6.8	1.4	3.6	13.1
	COL PCT	16.0	10.7	11.2	8.7	5.1	6.0	
	TOT PCT	9.2	.8	1.5	.9	.2	.5	
2 M CLASS NEXT DAY	COUNT	144	6	2	5	3	1	161
	ROW PCT	89.4	3.7	1.2	3.1	1.9	.6	3.6
	COL PCT	5.6	1.7	.3	1.1	1.9	.3	
	TOT PCT	3.2	.1	.0	.1	.1	.0	
3 X CLASS NEXT DAY	COUNT	18	3	1	0	0	0	22
	ROW PCT	81.8	13.6	4.5	0	0	0	.5
	COL PCT	.7	.8	.2	0	0	0	
	TOT PCT	.4	.1	.0	0	0	0	
COLUMN TOTAL		2569	355	597	458	156	352	4487
		57.3	7.9	13.3	10.2	3.5	7.8	100.0

RAW CHI SQUARE = 148.61877 WITH 15 DEGREES OF FREEDOM. SIGNIFICANCE = .0000
KENDALL#S TAU C = -.08872. SIGNIFICANCE = .0000
GAMMA = -.39971
SOMERS#S D (ASYMMETRIC) = -.10555 WITH FLARER DEPENDENT. = -.22555 WITH APPLONG DEPENDENT.
SOMERS#S D (SYMMETRIC) = -.14300

FLARER	CURLONG → COUNT / ROW PCT / COL PCT / TOT PCT	0 -120 (1)	121-150 (2)	151-180 (3)	181-210 (4)	211-240 (5)	241-270 (6)	OVER 270 (7)	ROW TOTAL
NO FLARE NXT DAY	0	523	702	470	327	731	633	332	3718
		14.1	18.9	12.6	8.8	19.7	17.0	8.9	82.9
		84.9	83.2	76.5	82.5	82.6	82.2	91.7	
		11.7	15.6	10.5	7.3	16.3	14.1	7.4	
C CLASS NEXT DAY	1	74	113	104	45	128	104	18	586
		12.6	19.3	17.7	7.7	21.8	17.7	3.1	13.1
		12.0	13.4	16.9	11.4	14.5	13.5	5.0	
		1.6	2.5	2.3	1.0	2.9	2.3	.4	
M CLASS NEXT DAY	2	18	25	37	20	23	27	11	161
		11.2	15.5	23.0	12.4	14.3	16.8	6.8	3.6
		2.9	3.0	6.0	5.1	2.6	3.5	3.0	
		.4	.6	.8	.4	.5	.6	.2	
X CLASS NEXT DAY	3	1	4	3	4	3	6	1	22
		4.5	18.2	13.6	18.2	13.6	27.3	4.5	.5
		.2	.5	.5	1.0	.3	.8	.3	
		.0	.1	.1	.1	.1	.1	.0	
	COLUMN TOTAL	616	844	614	396	885	770	362	4487
		13.7	18.8	13.7	8.8	19.7	17.2	8.1	100.0

RAW CHI SQUARE = 57.40781 WITH 18 DEGREES OF FREEDOM. SIGNIFICANCE = .0000
KENDALL'S TAU C = -.01202. SIGNIFICANCE = .0905
GAMMA = -.03625
SOMERS'S D (ASYMMETRIC) = -.01060 WITH FLARER DEPENDENT. = -.03057 WITH CURLONG DEPENDENT.
SOMERS'S D (SYMMETRIC) = -.01583

FLARER	NSLAT → COUNT / ROW PCT / COL PCT / TOT PCT	NORTH (1)	SOUTH (2)	ROW TOTAL
NO FLARE NXT DAY	0	2065	1652	3718
		55.6	44.4	82.9
		81.7	84.4	
		46.0	36.8	
C CLASS NEXT DAY	1	353	233	586
		60.2	39.8	13.1
		14.0	11.9	
		7.9	5.2	
M CLASS NEXT DAY	2	95	66	161
		59.0	41.0	3.6
		3.8	3.4	
		2.1	1.5	
X CLASS NEXT DAY	3	16	6	22
		72.7	27.3	.5
		.6	.3	
		.4	.1	
	COLUMN TOTAL	2530	1957	4487
		56.4	43.6	100.0

RAW CHI SQUARE = 7.38051 WITH 3 DEGREES OF FREEDOM. SIGNIFICANCE = .0605
KENDALL'S TAU C = -.02728. SIGNIFICANCE = .0074
GAMMA = -.09474
SOMERS'S D (ASYMMETRIC) = -.02774 WITH FLARER DEPENDENT. = -.04624 WITH NSLAT DEPENDENT.
SOMERS'S D (SYMMETRIC) = -.03467

CURLAT

FLARER	COUNT / ROW PCT / COL PCT / TOT PCT	0 -10 (1)	11-20 (2)	21-30 (3)	31-40 (4)	OVER 40 (5)	ROW TOTAL
0 NO FLARE NXT DAY	COUNT	110	1765	1472	341	30	3718
	ROW PCT	3.0	47.5	39.6	9.2	.8	82.9
	COL PCT	77.5	81.6	83.8	87.4	85.7	
	TOT PCT	2.5	39.3	32.8	7.6	.7	
1 C CLASS NEXT DAY	COUNT	23	289	228	42	4	586
	ROW PCT	3.9	49.3	38.9	7.2	.7	13.1
	COL PCT	16.2	13.4	13.0	10.8	11.4	
	TOT PCT	.5	6.4	5.1	.9	.1	
2 M CLASS NEXT DAY	COUNT	6	98	50	6	1	161
	ROW PCT	3.7	60.9	31.1	3.7	.6	3.6
	COL PCT	4.2	4.5	2.8	1.5	2.9	
	TOT PCT	.1	2.2	1.1	.1	.0	
3 X CLASS NEXT DAY	COUNT	3	11	7	1	0	22
	ROW PCT	13.6	50.0	31.8	4.5	0	.5
	COL PCT	2.1	.5	.4	.3	0	
	TOT PCT	.1	.2	.2	.0	0	
COLUMN TOTAL		142	2163	1757	390	35	4487
		3.2	48.2	39.2	8.7	.8	100.0

RAW CHI SQUARE = 26.36144 WITH 12 DEGREES OF FREEDOM. SIGNIFICANCE = .0095
KENDALL#S TAU C = -.02764. SIGNIFICANCE = .0002
GAMMA = -.11718
SOMERS#S D (ASYMMETRIC) = -.03423 WITH FLARER DEPENDENT. = -.07027 WITH CURLAT DEPENDENT.
SOMERS#S D (SYMMETRIC) = -.04604

CARLONG

FLARER	COUNT / ROW PCT / COL PCT / TOT PCT	0 -60 (1)	61-120 (2)	121-180 (3)	181-240 (4)	241-300 (5)	OVER 300 (6)	ROW TOTAL
0 NO FLARE NXT DAY	COUNT	555	654	747	647	537	578	3718
	ROW PCT	14.9	17.6	20.1	17.4	14.4	15.5	82.9
	COL PCT	79.5	82.0	84.6	82.0	84.0	85.0	
	TOT PCT	12.4	14.6	16.6	14.4	12.0	12.9	
1 C CLASS NEXT DAY	COUNT	100	104	105	116	86	75	586
	ROW PCT	17.1	17.7	17.9	19.8	14.7	12.8	13.1
	COL PCT	14.3	13.0	11.9	14.7	13.5	11.0	
	TOT PCT	2.2	2.3	2.3	2.6	1.9	1.7	
2 M CLASS NEXT DAY	COUNT	36	34	28	22	15	26	161
	ROW PCT	22.4	21.1	17.4	13.7	9.3	16.1	3.6
	COL PCT	5.2	4.3	3.2	2.8	2.3	3.8	
	TOT PCT	.8	.8	.6	.5	.3	.6	
3 X CLASS NEXT DAY	COUNT	7	6	3	4	1	1	22
	ROW PCT	31.8	27.3	13.6	18.2	4.5	4.5	.5
	COL PCT	1.0	.8	.3	.5	.2	.1	
	TOT PCT	.2	.1	.1	.1	.0	.0	
COLUMN TOTAL		698	798	883	789	639	680	4487
		15.6	17.8	19.7	17.6	14.2	15.2	100.0

RAW CHI SQUARE = 26.36375 WITH 15 DEGREES OF FREEDOM. SIGNIFICANCE = .0344
KENDALL#S TAU C = -.02312. SIGNIFICANCE = .0035
GAMMA = -.07066
SOMERS#S D (ASYMMETRIC) = -.02006 WITH FLARER DEPENDENT. = -.05878 WITH CARLONG DEPENDENT.
SOMERS#S D (SYMMETRIC) = -.03079

FLARER	COUNT / ROW PCT / COL PCT / TOT PCT	AGE 0	1	2	3	4	5	6	7	8	9	ROW TOTAL
0	COUNT	0	651	503	435	360	308	278	242	206	179	3718
NO FLARE NXT DAY	ROW PCT	0	17.5	13.5	11.7	9.7	8.3	7.5	6.5	5.5	4.8	82.9
	COL PCT	0	88.5	85.0	85.0	83.1	79.8	80.1	79.9	78.0	77.2	
	TOT PCT	0	14.5	11.2	9.7	8.0	6.9	6.2	5.4	4.6	4.0	
1	COUNT	1	65	73	60	57	64	53	42	44	45	586
C CLASS NEXT DAY	ROW PCT	.2	11.1	12.5	10.2	9.7	10.9	9.0	7.2	7.5	7.7	13.1
	COL PCT	100.0	8.8	12.3	11.7	13.2	16.6	15.3	13.9	16.7	19.4	
	TOT PCT	.0	1.4	1.6	1.3	1.3	1.4	1.2	.9	1.0	1.0	
2	COUNT	0	18	15	15	15	11	13	16	12	8	161
M CLASS NEXT DAY	ROW PCT	0	11.2	9.3	9.3	9.3	6.8	8.1	9.9	7.5	5.0	3.6
	COL PCT	0	2.4	2.5	2.9	3.5	2.8	3.7	5.3	4.5	3.4	
	TOT PCT	0	.4	.3	.3	.3	.2	.3	.4	.3	.2	
3	COUNT	0	2	1	2	1	3	3	3	2	0	22
X CLASS NEXT DAY	ROW PCT	0	9.1	4.5	9.1	4.5	13.6	13.6	13.6	9.1	0	.5
	COL PCT	0	.3	.2	.4	.2	.8	.9	1.0	.8	0	
	TOT PCT	0	.0	.0	.0	.0	.1	.1	.1	.0	0	
COLUMN TOTAL		1	736	592	512	433	386	347	303	264	232	4487
		.0	16.4	13.2	11.4	9.7	8.6	7.7	6.8	5.9	5.2	100.0

(CONTINUED)

FLARER	COUNT / ROW PCT / COL PCT / TOT PCT	AGE 10	11	12	13	14	15	ROW TOTAL
0	COUNT	164	147	115	92	34	4	3718
NO FLARE NXT DAY	ROW PCT	4.4	4.0	3.1	2.5	.9	.1	82.9
	COL PCT	78.8	79.9	80.4	88.5	91.9	80.0	
	TOT PCT	3.7	3.3	2.6	2.1	.8	.1	
1	COUNT	30	28	18	6	0	0	586
C CLASS NEXT DAY	ROW PCT	5.1	4.8	3.1	1.0	0	0	13.1
	COL PCT	14.4	15.2	12.6	5.8	0	0	
	TOT PCT	.7	.6	.4	.1	0	0	
2	COUNT	13	7	9	5	3	1	161
M CLASS NEXT DAY	ROW PCT	8.1	4.3	5.6	3.1	1.9	.6	3.6
	COL PCT	6.3	3.8	6.3	4.8	8.1	20.0	
	TOT PCT	.3	.2	.2	.1	.1	.0	
3	COUNT	1	2	1	1	0	0	22
X CLASS NEXT DAY	ROW PCT	4.5	9.1	4.5	4.5	0	0	.5
	COL PCT	.5	1.1	.7	1.0	0	0	
	TOT PCT	.0	.0	.0	.0	0	0	
COLUMN TOTAL		208	184	143	104	37	5	4487
		4.6	4.1	3.2	2.3	.8	.1	100.0

RAW CHI SQUARE = 82.53771 WITH 45 DEGREES OF FREEDOM. SIGNIFICANCE = .0005
KENDALL'S TAU C = .04406. SIGNIFICANCE = .0000
GAMMA = .12295
SOMERS'S D (ASYMMETRIC) = .03655 WITH FLARER DEPENDENT. = .11196 WITH AGE DEPENDENT.
SOMERS'S D (SYMMETRIC) = .05511

```
MAGCLAS
```

FLARER	COUNT / ROW PCT / COL PCT / TOT PCT	ALPHA 1	BETA 2	PETA-GAMMA 3	GAMMA 4	BETA-DELTA 5	BETA-GAMMA-DELTA 6	GAMMA-DELTA 7	9	ROW TOTAL
0	COUNT	1285	2188	198	8	12	10	2	144M	3704
NO FLARE NXT DAY	ROW PCT	34.7	59.1	5.3	.2	.3	.3	.1	0	82.9
	COL PCT	94.9	85.2	50.3	34.8	35.3	12.7	11.8	0	
	TOT PCT	28.8	48.9	4.4	.2	.3	.2	.0	0	
1	COUNT	56	318	149	5	8	39	7	2M	584
C CLASS NEXT DAY	ROW PCT	9.9	54.5	25.5	.9	1.4	6.7	1.2	0	13.1
	COL PCT	4.3	12.4	37.8	21.7	23.5	49.4	41.2	0	
	TOT PCT	1.3	7.1	3.3	.1	.2	.9	.2	0	
2	COUNT	10	58	44	9	10	25	4	19	160
M CLASS NEXT DAY	ROW PCT	6.3	36.2	27.5	5.6	6.3	15.6	2.5	0	3.6
	COL PCT	.7	2.3	11.2	39.1	29.4	31.6	23.5	0	
	TOT PCT	.2	1.3	1.0	.2	.2	.6	.1	0	
3	COUNT	1	4	3	1	4	5	4	0M	22
X CLASS NEXT DAY	ROW PCT	4.5	18.2	13.6	4.5	18.2	22.7	18.2	0	.5
	COL PCT	.1	.2	.8	4.3	11.8	6.3	23.5	0	
	TOT PCT	.0	.1	.1	.0	.1	.1	.1	0	
COLUMN TOTAL	COUNT	1355	2568	394	23	34	79	17	17M	4470
	PCT	30.3	57.4	8.8	.5	.8	1.8	.4	0	100.0

```
RAW CHI SQUARE =  1296.74344 WITH    18 DEGREES OF FREEDOM.  SIGNIFICANCE =       0
KENDALL#S TAU C =   .18521.     SIGNIFICANCE =       0
GAMMA =   .70888
SOMERS#S D (ASYMMETRIC) =    .24375 WITH FLARER   DEPENDENT.          =   .47089 WITH MAGCLAS  DEPENDENT.
SOMERS#S D (SYMMETRIC) =   .32122

NUMBER OF MISSING OBSERVATIONS =    17
```

```
RV
```

FLARER	COUNT / ROW PCT / COL PCT / TOT PCT	NO SPOTS 0	RED 1	VIOLET 2	NO DATA 9	ROW TOTAL
0	COUNT	27	893	713	2095M	1633
NO FLARE NXT DAY	ROW PCT	1.7	54.7	43.7	0	80.3
	COL PCT	87.1	80.9	79.3	0	
	TOT PCT	1.3	43.9	35.1	0	
1	COUNT	4	144	148	290M	296
C CLASS NEXT DAY	ROW PCT	1.4	48.6	50.0	0	14.6
	COL PCT	12.9	13.0	16.5	0	
	TOT PCT	.2	7.1	7.3	0	
2	COUNT	0	57	31	73M	88
M CLASS NEXT DAY	ROW PCT	0	64.8	35.2	0	4.3
	COL PCT	0	5.2	3.4	0	
	TOT PCT	0	2.8	1.5	0	
3	COUNT	0	10	7	5M	17
X CLASS NEXT DAY	ROW PCT	0	58.8	41.2	0	.8
	COL PCT	0	.9	.8	0	
	TOT PCT	0	.5	.3	0	
COLUMN TOTAL	COUNT	31	1104	899	2453M	2034
	PCT	1.5	54.3	44.2	0	100.0

```
RAW CHI SQUARE =     9.46378 WITH     6 DEGREES OF FREEDOM.  SIGNIFICANCE =   .1491
KENDALL#S TAU C =   .01161.     SIGNIFICANCE =   .1918
GAMMA =   .04573
SOMERS#S D (ASYMMETRIC) =    .01519 WITH FLARER   DEPENDENT.          =   .02338 WITH RV        DEPENDENT.
SOMERS#S D (SYMMETRIC) =   .01839

NUMBER OF MISSING OBSERVATIONS =  2453
```

MAGSTR

FLARER	COUNT / ROW PCT / COL PCT / TOT PCT	0 - 10 (101)	11- 20 (102)	OVER 20 (103)	NO DATA (99)	ROW TOTAL
NO FLARE NXT DAY	0	285	1159	178	2095M	1623
		17.5	71.4	11.0	0	80.3
		92.6	80.8	64.0	0	
		14.1	57.3	8.8	0	
C CLASS NEXT DAY	1	20	217	57	2929M	294
		6.8	73.8	19.4	0	14.5
		6.5	15.1	20.5	0	
		1.0	10.7	2.8	0	
M CLASS NEXT DAY	2	3	50	35	73M	88
		3.4	56.8	39.8	0	4.4
		1.0	3.5	12.6	0	
		.1	2.5	1.7	0	
X CLASS NEXT DAY	3	0	9	8	5M	17
		0	52.9	47.1	0	.8
		0	.6	2.9	0	
		0	.4	.4	0	
COLUMN TOTAL		309	1435	278	2465M	2022
		15.3	71.0	13.7	0	100.0

RAW CHI SQUARE = 107.00091 WITH 6 DEGREES OF FREEDOM. SIGNIFICANCE = .0000
KENDALL'S TAU C = .11071. SIGNIFICANCE = .0000
GAMMA = .47203
SOMERS'S D (ASYMMETRIC) = .16254 WITH FLARER DEPENDENT. = .22190 WITH MAGSTR DEPENDENT.
SOMERS'S D (SYMMETRIC) = .18764

NUMBER OF MISSING OBSERVATIONS = 2465

MAGGRAD

FLARER	COUNT / ROW PCT / COL PCT / TOT PCT	0.00 (100)	.01-.10 (101)	.11-.20 (102)	.21-.30 (103)	.31-.99 (104)	NO DATA (9)	ROW TOTAL
NO FLARE NXT DAY	0	1217	613	150	19	6	1713M	2005
		60.7	30.6	7.5	.9	.3	0	84.0
		94.9	80.2	57.0	33.9	30.0	0	
		51.0	25.7	6.3	.8	.3	0	
C CLASS NEXT DAY	1	56	123	81	19	6	301M	285
		19.6	43.2	28.4	6.7	2.1	0	11.9
		4.4	16.1	30.8	33.9	30.0	0	
		2.3	5.2	3.4	.8	.3	0	
M CLASS NEXT DAY	2	10	25	26	14	6	80M	81
		12.3	30.9	32.1	17.3	7.4	0	3.4
		.8	3.3	9.9	25.0	30.0	0	
		.4	1.0	1.1	.6	.3	0	
X CLASS NEXT DAY	3	0	3	6	4	2	7M	15
		0	20.0	40.0	26.7	13.3	0	.6
		0	.4	2.3	7.1	10.0	0	
		0	.1	.3	.2	.1	0	
COLUMN TOTAL		1283	764	263	56	20	2181M	2386
		53.8	32.0	11.0	2.3	.8	0	100.0

RAW CHI SQUARE = 507.96255 WITH 12 DEGREES OF FREEDOM. SIGNIFICANCE = 0
KENDALL'S TAU C = .19613. SIGNIFICANCE = 0
GAMMA = .70423
SOMERS'S D (ASYMMETRIC) = .24699 WITH FLARER DEPENDENT. = .52836 WITH MAGGRAD DEPENDENT.
SOMERS'S D (SYMMETRIC) = .33663

NUMBER OF MISSING OBSERVATIONS = 2181

SSDYNAM

FLARER	COUNT / ROW PCT / COL PCT / TOT PCT	NO SPOTS OR NA 0	COALESCING 1	SPOT ROTATION 2	RELATIVE MOTION 3	NO DATA 9	ROW TOTAL
0 NO FLARE NXT DAY	COUNT	3608	33	7	10	60M	3658
	ROW PCT	98.6	.9	.2	.3	0	82.9
	COL PCT	83.9	50.0	43.8	29.4	0	
	TOT PCT	81.7	.7	.2	.2	0	
1 C CLASS NEXT DAY	COUNT	534	23	6	17	6M	580
	ROW PCT	92.1	4.0	1.0	2.9	0	13.1
	COL PCT	12.4	34.8	37.5	50.0	0	
	TOT PCT	12.1	.5	.1	.4	0	
2 M CLASS NEXT DAY	COUNT	140	9	2	4	6M	155
	ROW PCT	90.3	5.8	1.3	2.6	0	3.5
	COL PCT	3.3	13.6	12.5	11.8	0	
	TOT PCT	3.2	.2	.0	.1	0	
3 X CLASS NEXT DAY	COUNT	16	1	1	3	1M	21
	ROW PCT	76.2	4.8	4.8	14.3	0	.5
	COL PCT	.4	1.5	6.3	8.8	0	
	TOT PCT	.4	.0	.0	.1	0	
COLUMN TOTAL		4298	66	16	34	73M	4414
		97.4	1.5	.4	.8	0	100.0

RAW CHI SQUARE = 185.42537 WITH 9 DEGREES OF FREEDOM. SIGNIFICANCE = 0
KENDALL#S TAU C = .02850. SIGNIFICANCE = .0000
GAMMA = .70839
SOMERS#S D (ASYMMETRIC) = .41444 WITH FLARER DEPENDENT. = .07253 WITH SSDYNAM DEPENDENT.
SOMERS#S D (SYMMETRIC) = .12346

NUMBER OF MISSING OBSERVATIONS = 73

SSINTER

FLARER	COUNT / ROW PCT / COL PCT / TOT PCT	NO INTERACTION 0	SPOTS CONVERGE 1	NO DATA 9	ROW TOTAL
0 NO FLARE NXT DAY	COUNT	3653	27	38M	3680
	ROW PCT	99.3	.7	0	82.9
	COL PCT	83.2	58.7	0	
	TOT PCT	82.3	.6	0	
1 C CLASS NEXT DAY	COUNT	569	13	4M	582
	ROW PCT	97.8	2.2	0	13.1
	COL PCT	13.0	28.3	0	
	TOT PCT	12.8	.3	0	
2 M CLASS NEXT DAY	COUNT	151	4	6M	155
	ROW PCT	97.4	2.6	0	3.5
	COL PCT	3.4	8.7	0	
	TOT PCT	3.4	.1	0	
3 X CLASS NEXT DAY	COUNT	19	2	1M	21
	ROW PCT	90.5	9.5	0	.5
	COL PCT	.4	4.3	0	
	TOT PCT	.4	.0	0	
COLUMN TOTAL		4392	46	49M	4438
		99.0	1.0	0	100.0

RAW CHI SQUARE = 25.77174 WITH 3 DEGREES OF FREEDOM. SIGNIFICANCE = .0000
KENDALL#S TAU C = .01033. SIGNIFICANCE = .0000
GAMMA = .53359
SOMERS#S D (ASYMMETRIC) = .25186 WITH FLARER DEPENDENT. = .01758 WITH SSINTER DEPENDENT.
SOMERS#S D (SYMMETRIC) = .03286

NUMBER OF MISSING OBSERVATIONS = 49

STGDEV

FLARER	COUNT / ROW PCT / COL PCT / TOT PCT	IMATURE GROUP 1	DECAYING 2	GROWING 3	RAPID DECAY 4	RAPID GROWTH 5	EXTREME GROWTH 6	NO DATA 9	ROW TOTAL
NO FLARE NXT DAY	0	1733	743	912	45	42	40	203M	3515
		49.3	21.1	25.9	1.3	1.2	1.1	0	82.7
		86.2	85.1	76.6	77.6	66.7	75.5	0	
		40.8	17.5	21.5	1.1	1.0	.9	0	
C CLASS NEXT DAY	1	214	102	210	8	17	10	25M	561
		38.1	18.2	37.4	1.4	3.0	1.8	0	13.2
		10.5	11.7	17.6	13.8	27.0	18.9	0	
		5.0	2.4	4.9	.2	.4	.2	0	
M CLASS NEXT DAY	2	59	22	62	3	4	3	8M	153
		38.6	14.4	40.5	2.0	2.6	2.0	0	3.6
		2.9	2.5	5.2	5.2	6.3	5.7	0	
		1.4	.5	1.5	.1	.1	.1	0	
X CLASS NEXT DAY	3	5	6	7	2	0	0	2M	20
		25.0	30.0	35.0	10.0	0	0	0	.5
		.2	.7	.6	3.4	0	0	0	
		.1	.1	.2	.0	0	0	0	
	COLUMN TOTAL	2011	873	1191	58	63	53	238M	4249
		47.3	20.5	28.0	1.4	1.5	1.2	0	100.0

RAW CHI SQUARE = 81.98311 WITH 15 DEGREES OF FREEDOM. SIGNIFICANCE = .0000
KENDALL'S TAU C = .06091. SIGNIFICANCE = .0000
GAMMA = .22742
SOMERS'S D (ASYMMETRIC) = .06978 WITH FLARER DEPENDENT. = .15387 WITH STGDEV DEPENDENT.
SOMERS'S D (SYMMETRIC) = .09602

NUMBER OF MISSING OBSERVATIONS = 238

C.3 H-Alpha Variables

LEADTRAI

FLARES	COUNT / ROW PCT / COL PCT / TOT PCT	STRUCT NOT DEF 0	RETURN REGION 1	< 5 NL OUT PHAS 2	> 5 NL LEADER 3	> 5 NL TRAILER 4	< 5 NL IN PHASE 5	ROW TOTAL
0	COUNT	40	1015	567	809	757	530	3718
NO FLARE NXT DAY	ROW PCT	1.1	27.3	15.3	21.8	20.4	14.3	82.9
	COL PCT	97.6	78.3	84.9	86.4	82.7	84.1	
	TOT PCT	.9	22.6	12.6	18.0	16.9	11.8	
1	COUNT	1	200	86	96	121	82	586
C CLASS NEXT DAY	ROW PCT	.2	34.1	14.7	16.4	20.6	14.0	13.1
	COL PCT	2.4	15.4	12.9	10.3	13.2	13.0	
	TOT PCT	.0	4.5	1.9	2.1	2.7	1.8	
2	COUNT	0	71	14	30	30	16	161
M CLASS NEXT DAY	ROW PCT	0	44.1	8.7	18.6	18.6	9.9	3.6
	COL PCT	0	5.5	2.1	3.2	3.3	2.5	
	TOT PCT	0	1.6	.3	.7	.7	.4	
3	COUNT	0	11	1	1	7	2	22
X CLASS NEXT DAY	ROW PCT	0	50.0	4.5	4.5	31.8	9.1	.5
	COL PCT	0	.8	.1	.1	.8	.3	
	TOT PCT	0	.2	.0	.0	.2	.0	
	COLUMN TOTAL	41	1297	668	936	915	630	4487
		.9	28.9	14.9	20.9	20.4	14.0	100.0

RAW CHI SQUARE = 51.92284 WITH 15 DEGREES OF FREEDOM. SIGNIFICANCE = .0000
KENDALL'S TAU C = -.02608. SIGNIFICANCE = .0011
GAMMA = -.08476
SOMERS'S D (ASYMMETRIC) = -.02478 WITH FLARES DEPENDENT. = -.06630 WITH LEADTRAI DEPENDENT.
SOMERS'S D (SYMMETRIC) = -.03608

RETREG

FLARES	COUNT / ROW PCT / COL PCT / TOT PCT	NOT RET 0	RETURN 1	ROW TOTAL
0	COUNT	2703	1015	3718
NO FLARE NXT DAY	ROW PCT	72.7	27.3	82.9
	COL PCT	84.7	78.3	
	TOT PCT	60.2	22.6	
1	COUNT	386	200	586
C CLASS NEXT DAY	ROW PCT	65.9	34.1	13.1
	COL PCT	12.1	15.4	
	TOT PCT	8.5	4.5	
2	COUNT	90	71	161
M CLASS NEXT DAY	ROW PCT	55.9	44.1	3.6
	COL PCT	2.8	5.5	
	TOT PCT	2.0	1.6	
3	COUNT	11	11	22
X CLASS NEXT DAY	ROW PCT	50.0	50.0	.5
	COL PCT	.3	.8	
	TOT PCT	.2	.2	
	COLUMN TOTAL	3190	1297	4487
		71.1	28.9	100.0

RAW CHI SQUARE = 35.29792 WITH 3 DEGREES OF FREEDOM. SIGNIFICANCE = .0000
KENDALL'S TAU C = .05555. SIGNIFICANCE = .0000
GAMMA = .21341
SOMERS'S D (ASYMMETRIC) = .06758 WITH FLARES DEPENDENT. = .09414 WITH RETREG DEPENDENT.
SOMERS'S D (SYMMETRIC) = .07868

SECTEON

FLARER	COUNT ROW PCT COL PCT TOT PCT	STRUCT NOT DEF 0	>30 FRM BOUNDARY 1	NON-HALE 10-30 W 2	NON-HALE 10-30 E 3	NON-HALE <10 BND 4	HALE 10-30 W 5	HALE 10-30 E 6	HALE <10 BND 7	NO DATA 9	ROW TOTAL
NO FLARE NXT DAY	0	1 .0 100.0 .0	1622 43.6 83.4 36.2	379 10.2 91.5 8.4	247 6.6 85.2 5.5	272 7.3 76.8 6.1	403 10.8 81.9 9.0	332 8.9 79.2 7.4	461 12.4 80.9 10.3	1M 0 0 0	3717 82.9
C CLASS NEXT DAY	1	0 0 0 0	240 41.0 12.3 5.3	34 5.8 8.2 .8	37 6.3 12.8 .8	67 11.4 18.9 1.5	70 11.9 14.2 1.6	64 10.9 15.3 1.4	74 12.6 13.0 1.6	0M 0 0 0	566 13.1
M CLASS NEXT DAY	2	0 0 0 0	75 46.6 3.9 1.7	1 .6 .2 .0	5 3.1 1.7 .1	15 9.3 4.2 .3	17 10.6 3.5 .4	20 12.4 4.8 .4	28 17.4 4.9 .6	0M 0 0 0	161 3.6
X CLASS NEXT DAY	3	0 0 0 0	9 40.9 .5 .2	0 0 0 0	1 4.5 .3 .0	0 0 0 0	2 9.1 .4 .0	3 13.6 .7 .1	7 31.8 1.2 .2	0M 0 0 0	22 .5
COLUMN TOTAL		1 .0	1946 43.4	414 9.2	290 6.5	354 7.9	492 11.0	419 9.3	570 12.7	1M 0	4486 100.0

RAW CHI SQUARE = 58.15583 WITH 21 DEGREES OF FREEDOM. SIGNIFICANCE = .0000
KENDALL'S TAU C = .02053. SIGNIFICANCE = .0067
GAMMA = .06871
SOMERS'S D (ASYMMETRIC) = .02037 WITH FLARER DEPENDENT. = .05218 WITH SECTEON DEPENDENT.
SOMERS'S D (SYMMETRIC) = .02930

NUMBER OF MISSING OBSERVATIONS = 1

PLAGFIL

FLARER	COUNT ROW PCT COL PCT TOT PCT	NON-COM NO FILMT 0	NON-COM WI FILMT 1	NON-COM ACT FIL 2	COMPACT NO FILMT 3	COMPACT WI FILMT 4	COMPACT ACT FIL 5	6	NO DATA 9	ROW TOTAL
NO FLARE NXT DAY	0	1654 50.5 90.3 41.5	674 20.6 87.5 16.9	130 4.0 69.5 3.3	640 19.5 71.5 16.1	133 4.1 68.2 3.3	47 1.4 48.0 1.2	0 0 0 0	440M 0 0 0	3278 82.4
C CLASS NEXT DAY	1	151 27.8 8.2 3.8	75 13.8 9.7 1.9	42 7.7 22.5 1.1	194 35.7 21.7 4.9	47 8.6 24.1 1.2	35 6.4 35.7 .9	0 0 0 0	42M 0 0 0	544 13.7
M CLASS NEXT DAY	2	24 17.8 1.3 .6	19 14.1 2.5 .5	11 8.1 5.9 .3	54 40.0 6.0 1.4	13 9.6 6.7 .3	14 10.4 14.3 .4	0 0 0 0	26M 0 0 0	135 3.4
X CLASS NEXT DAY	3	2 10.0 .1 .1	2 10.0 .3 .1	4 20.0 2.1 .1	7 35.0 .8 .2	2 10.0 1.0 .1	2 10.0 2.0 .1	1 5.0 100.0 .0	2M 0 0 0	20 .5
COLUMN TOTAL		1831 46.0	770 19.4	187 4.7	895 22.5	195 4.9	98 2.5	1 .0	510M 0	3977 100.0

RAW CHI SQUARE = 511.79580 WITH 18 DEGREES OF FREEDOM. SIGNIFICANCE = 0
KENDALL'S TAU C = .13765. SIGNIFICANCE = 0
GAMMA = .44067
SOMERS'S D (ASYMMETRIC) = .14061 WITH FLARER DEPENDENT. = .34327 WITH PLAGFIL DEPENDENT.
SOMERS'S D (SYMMETRIC) = .20742

NUMBER OF MISSING OBSERVATIONS = 510

FLARER	NEUTLOR COUNT ROW PCT COL PCT TOT PCT	WEAK STRUCT 0	NORTH SOUTH 1	EAST WEST 2	HAIRPIN E-W 3	MOSTLY CIRCULAR 4	NO DATA 9	ROW TOTAL
NO FLARE NXT DAY	0	906	2228	292	60	166	66M	3652
		24.8	61.0	8.0	1.6	4.5	0	83.0
		89.2	84.5	73.0	53.1	70.6	0	
		20.5	50.6	6.6	1.4	3.8	0	
C CLASS NEXT DAY	1	74	329	81	36	53	13M	573
		12.9	57.4	14.1	6.3	9.2	0	13.0
		7.3	12.5	20.2	31.9	22.6	0	
		1.7	7.5	1.8	.8	1.2	0	
M CLASS NEXT DAY	2	34	68	23	14	15	7M	154
		22.1	44.2	14.9	9.1	9.7	0	3.5
		3.3	2.6	5.7	12.4	6.4	0	
		.8	1.5	.5	.3	.3	0	
X CLASS NEXT DAY	3	2	11	4	3	1	1M	21
		9.5	52.4	19.0	14.3	4.8	0	.5
		.2	.4	1.0	2.7	.4	0	
		.0	.2	.1	.1	.0	0	
	COLUMN TOTAL	1015	2636	400	113	235	87M	4400
		23.1	59.9	9.1	2.6	5.3	0	100.0

RAW CHI SQUARE = 174.96110 WITH 12 DEGREES OF FREEDOM. SIGNIFICANCE = 0
KENDALL'S TAU C = .08324. SIGNIFICANCE = .0000
GAMMA = .33343
SOMERS'S D (ASYMMETRIC) = .10448 WITH FLARER DEPENDENT. = .20546 WITH NEUTLOR DEPENDENT.
SOMERS'S D (SYMMETRIC) = .13852

NUMBER OF MISSING OBSERVATIONS = 87

FLARER	REVPOL COUNT ROW PCT COL PCT TOT PCT	NO REVERSE 0	REVERSE POLARITY 1	ROW TOTAL
NO FLARE NXT DAY	0	3708	10	3718
		99.7	.3	82.9
		83.0	50.0	
		82.6	.2	
C CLASS NEXT DAY	1	578	8	586
		98.6	1.4	13.1
		12.9	40.0	
		12.9	.2	
M CLASS NEXT DAY	2	160	1	161
		99.4	.6	3.6
		3.6	5.0	
		3.6	.0	
X CLASS NEXT DAY	3	21	1	22
		95.5	4.5	.5
		.5	5.0	
		.5	.0	
	COLUMN TOTAL	4467	20	4487
		99.6	.4	100.0

RAW CHI SQUARE = 22.22682 WITH 3 DEGREES OF FREEDOM. SIGNIFICANCE = .0001
KENDALL'S TAU C = .00503. SIGNIFICANCE = .0001
GAMMA = .61821
SOMERS'S D (ASYMMETRIC) = .32037 WITH FLARER DEPENDENT. = .00908 WITH REVPOL DEPENDENT.
SOMERS'S D (SYMMETRIC) = .01918

NEUTLCOM

FLARER	COUNT / ROW PCT / COL PCT / TOT PCT	NO KINKS WK STRC 0	1 - 3 KINKS 1	4 - 6 KINKS 2	7 - 12 KINKS 3	> 12 INKS 4	K NO DATA 9	ROW TOTAL
NO FLARE NXT DAY	0	1168	1457	395	125	13	560M	3158
		37.0	46.1	12.5	4.0	.4	0	82.3
		93.1	85.7	68.5	49.2	26.5	0	
		30.5	38.0	10.3	3.3	.3	0	
C CLASS NEXT DAY	1	69	206	135	96	16	64M	522
		13.2	39.5	25.9	18.4	3.1	0	13.6
		5.5	12.1	23.4	37.8	32.7	0	
		1.8	5.4	3.5	2.5	.4	0	
M CLASS NEXT DAY	2	17	35	43	25	16	25M	136
		12.5	25.7	31.6	18.4	11.8	0	3.5
		1.4	2.1	7.5	9.8	32.7	0	
		.4	.9	1.1	.7	.4	0	
X CLASS NEXT DAY	3	0	3	4	8	4	3M	19
		0	15.8	21.1	42.1	21.1	0	.5
		0	.2	.7	3.1	8.2	0	
		0	.1	.1	.2	.1	0	
	COLUMN TOTAL	1254	1701	577	254	49	652M	3835
		32.7	44.4	15.0	6.6	1.3	0	100.0

RAW CHI SQUARE = 614.41364 WITH 12 DEGREES OF FREEDOM. SIGNIFICANCE = 0
KENDALL'S TAU C = .17313. SIGNIFICANCE = 0
GAMMA = .57602
SOMERS'S D (ASYMMETRIC) = .19404 WITH FLARER DEPENDENT. = .42902 WITH NEUTLCOM DEPENDENT.
SOMERS'S D (SYMMETRIC) = .26738

NUMBER OF MISSING OBSERVATIONS = 652

NEUTLCHG

FLARER	COUNT / ROW PCT / COL PCT / TOT PCT	NO DEF TREND 0	SIMPLER 1	MORE COMPLEX 2	NO DATA 9	ROW TOTAL
NO FLARE NXT DAY	0	2550	185	232	751M	2967
		85.9	6.2	7.8	0	82.2
		85.2	83.0	58.9	0	
		70.5	5.1	6.4	0	
C CLASS NEXT DAY	1	355	30	119	82M	504
		70.4	6.0	23.6	0	14.0
		11.9	13.5	30.2	0	
		9.8	.8	3.3	0	
M CLASS NEXT DAY	2	76	7	38	40M	121
		62.8	5.8	31.4	0	3.4
		2.5	3.1	9.6	0	
		2.1	.2	1.1	0	
X CLASS NEXT DAY	3	13	1	5	3M	19
		68.4	5.3	26.3	0	.5
		.4	.4	1.3	0	
		.4	.0	.1	0	
	COLUMN TOTAL	2994	223	394	876M	3611
		82.9	6.2	10.9	0	100.0

RAW CHI SQUARE = 170.44896 WITH 6 DEGREES OF FREEDOM. SIGNIFICANCE = 0
KENDALL'S TAU C = .08077. SIGNIFICANCE = .0000
GAMMA = .45671
SOMERS'S D (ASYMMETRIC) = .10141 WITH FLARER DEPENDENT. = .17698 WITH NEUTLCHG DEPENDENT.
SOMERS'S D (SYMMETRIC) = .17017

NUMBER OF MISSING OBSERVATIONS = 876

ASSOCFIL

FLARER	COUNT / ROW PCT / COL PCT / TOT PCT	NO FILAMENT 0	FILAMENT UNCHGD 1	FILAMENT GROWING 2	FILAMENT DISAPPD 3	DARKENS OR ACTIV 4	NO DATA 9	ROW TOTAL
0 NO FLARE NXT DAY	COUNT	1819	809	169	66	131	724M	2994
	ROW PCT	60.8	27.0	5.6	2.2	4.4	0	82.2
	COL PCT	84.0	83.1	75.1	79.5	67.9	0	
	TOT PCT	50.0	22.2	4.6	1.8	3.6	0	
1 C CLASS NEXT DAY	COUNT	287	120	42	14	44	79M	507
	ROW PCT	56.6	23.7	8.3	2.8	8.7	0	13.9
	COL PCT	13.3	12.3	18.7	16.9	22.8	0	
	TOT PCT	7.9	3.3	1.2	.4	1.2	0	
2 M CLASS NEXT DAY	COUNT	53	38	13	2	15	40M	121
	ROW PCT	43.8	31.4	10.7	1.7	12.4	0	3.3
	COL PCT	2.4	3.9	5.8	2.4	7.8	0	
	TOT PCT	1.5	1.0	.4	.1	.4	0	
3 X CLASS NEXT DAY	COUNT	7	7	1	1	3	3M	19
	ROW PCT	36.8	36.8	5.3	5.3	15.8	0	.5
	COL PCT	.3	.7	.4	1.2	1.6	0	
	TOT PCT	.2	.2	.0	.0	.1	0	
COLUMN TOTAL		2166	974	225	83	193	846M	3641
		59.5	26.8	6.2	2.3	5.3	0	100.0

RAW CHI SQUARE = 53.59445 WITH 12 DEGREES OF FREEDOM. SIGNIFICANCE = .0000
KENDALL#S TAU C = .04039. SIGNIFICANCE = .0000
GAMMA = .16702
SOMERS#S D (ASYMMETRIC) = .05339 WITH FLARER DEPENDENT. = .09968 WITH ASSOCFIL DEPENDENT.
SOMERS#S D (SYMMETRIC) = .06959

NUMBER OF MISSING OBSERVATIONS = 846

BRTPTS

FLARER	COUNT / ROW PCT / COL PCT / TOT PCT	NO BRT PTS 0	NOT ALONG NL 1	ALONG NL 2	NO DATA 9	ROW TOTAL
0 NO FLARE NXT DAY	COUNT	2941	432	291	54M	3664
	ROW PCT	80.3	11.8	7.9	0	82.9
	COL PCT	89.5	68.9	57.4	0	
	TOT PCT	66.5	9.8	6.6	0	
1 C CLASS NEXT DAY	COUNT	274	143	163	6M	580
	ROW PCT	47.2	24.7	28.1	0	13.1
	COL PCT	8.3	22.8	32.1	0	
	TOT PCT	6.2	3.2	3.7	0	
2 M CLASS NEXT DAY	COUNT	65	44	47	5M	156
	ROW PCT	41.7	28.2	30.1	0	3.5
	COL PCT	2.0	7.0	9.3	0	
	TOT PCT	1.5	1.0	1.1	0	
3 X CLASS NEXT DAY	COUNT	7	8	6	1M	21
	ROW PCT	33.3	38.1	28.6	0	.5
	COL PCT	.2	1.3	1.2	0	
	TOT PCT	.2	.2	.1	0	
COLUMN TOTAL		3287	627	507	66M	4421
		74.3	14.2	11.5	0	100.0

RAW CHI SQUARE = 423.88523 WITH 6 DEGREES OF FREEDOM. SIGNIFICANCE = 0
KENDALL#S TAU C = .15367. SIGNIFICANCE = 0
GAMMA = .39664
SOMERS#S D (ASYMMETRIC) = .24760 WITH FLARER DEPENDENT. = .34760 WITH BRTPTS DEPENDENT.
SOMERS#S D (SYMMETRIC) = .28915

NUMBER OF MISSING OBSERVATIONS = 66

PLAGFLUX

FLARER	COUNT / ROW PCT / COL PCT / TOT PCT	NO PLAGE FLUCTS 0	PLAGE FLUCTS 1	NO DATA 9	ROW TOTAL
NO FLARE NXT DAY	0	3167	497	54M	3664
		86.4	13.6	0	82.9
		85.9	68.2	0	
		71.5	11.2	0	
C CLASS NEXT DAY	1	412	168	6M	580
		71.0	29.0	0	13.1
		11.2	23.0	0	
		9.3	3.8	0	
M CLASS NEXT DAY	2	101	55	5M	156
		64.7	35.3	0	3.5
		2.7	7.5	0	
		2.3	1.2	0	
X CLASS NEXT DAY	3	12	9	1M	21
		57.1	42.9	0	.5
		.3	1.2	0	
		.3	.2	0	
	COLUMN TOTAL	3692	729	66M	4421
		83.5	16.5	0	100.0

RAW CHI SQUARE = 138.82648 WITH 3 DEGREES OF FREEDOM. SIGNIFICANCE = 0
KENDALL#S TAU C = .09853. SIGNIFICANCE = .0000
GAMMA = .46178
SOMERS#S D (ASYMMETRIC) = .17888 WITH FLARER DEPENDENT. = .16719 WITH PLAGFLUX DEPENDENT.
SOMERS#S D (SYMMETRIC) = .17284

NUMBER OF MISSING OBSERVATIONS = 66

ISOPOLE

FLARER	COUNT / ROW PCT / COL PCT / TOT PCT	NO ISOL POLE 0	ISOLATED POLE 1	NO DATA 9	ROW TOTAL
NO FLARE NXT DAY	0	3619	45	54M	3664
		98.8	1.2	0	82.9
		83.7	46.4	0	
		81.9	1.0	0	
C CLASS NEXT DAY	1	544	35	7M	579
		94.0	6.0	0	13.1
		12.5	36.1	0	
		12.3	.8	0	
M CLASS NEXT DAY	2	138	17	6M	155
		89.0	11.0	0	3.5
		3.2	17.5	0	
		3.1	.4	0	
X CLASS NEXT DAY	3	21	0	1M	21
		100.0	0	0	.5
		.5	0	0	
		.5	0	0	
	COLUMN TOTAL	4322	97	68M	4419
		97.8	2.2	0	100.0

RAW CHI SQUARE = 111.96195 WITH 3 DEGREES OF FREEDOM. SIGNIFICANCE = .0000
KENDALL#S TAU C = .03275. SIGNIFICANCE = .0000
GAMMA = .68036
SOMERS#S D (ASYMMETRIC) = .30136 WITH FLARER DEPENDENT. = .05560 WITH ISOPOLE DEPENDENT.
SOMERS#S D (SYMMETRIC) = .09717

NUMBER OF MISSING OBSERVATIONS = 68

FLARER	COUNT / ROW PCT / COL PCT / TOT PCT	NONE OR NEW 0	EMERGES IN GROUP 1	EMERGES NEAR REG 2	NO DATA 9	ROW TOTAL
NO FLARE NXT DAY	0	3587	57	20	54M	3664
		97.9	1.6	.5	0	82.9
		83.4	66.3	64.5	0	
		81.2	1.3	.5	0	
C CLASS NEXT DAY	1	547	24	8	7M	579
		94.5	4.1	1.4	0	13.1
		12.7	27.9	25.8	0	
		12.4	.5	.2	0	
M CLASS NEXT DAY	2	148	4	3	6M	155
		95.5	2.6	1.9	0	3.5
		3.4	4.7	9.7	0	
		3.3	.1	.1	0	
X CLASS NEXT DAY	3	20	1	0	1M	21
		95.2	4.8	0	0	.5
		.5	1.2	0	0	
		.5	.0	0	0	
	COLUMN TOTAL	4302	86	31	68M	4419
		97.4	1.9	.7	0	100.0

(Column variable: EFR)

RAW CHI SQUARE = 27.69842 WITH 6 DEGREES OF FREEDOM. SIGNIFICANCE = .0001
KENDALL'S TAU C = .01344. SIGNIFICANCE = .0000
GAMMA = .41691
SOMERS'S D (ASYMMETRIC) = .17294 WITH FLARER DEPENDENT. = .03848 WITH EFR DEPENDENT.
SOMERS'S D (SYMMETRIC) = .05182

NUMBER OF MISSING OBSERVATIONS = 68

FLARER	COUNT / ROW PCT / COL PCT / TOT PCT	NO AFS 0	AFS PRESENT 1	NO DATA 9	ROW TOTAL
NO FLARE NXT DAY	0	3169	496	54M	3664
		86.5	13.5	0	82.9
		85.3	70.2	0	
		71.7	11.2	0	
C CLASS NEXT DAY	1	406	173	7M	579
		70.1	29.9	0	13.1
		10.9	24.5	0	
		9.2	3.9	0	
M CLASS NEXT DAY	2	123	35	6M	155
		77.4	22.6	0	3.5
		3.2	5.0	0	
		2.7	.8	0	
X CLASS NEXT DAY	3	18	3	1M	21
		85.7	14.3	0	.5
		.5	.4	0	
		.4	.1	0	
	COLUMN TOTAL	3712	707	68M	4419
		84.0	16.0	0	100.0

(Column variable: AFS)

RAW CHI SQUARE = 104.56690 WITH 3 DEGREES OF FREEDOM. SIGNIFICANCE = .0000
KENDALL'S TAU C = .07987. SIGNIFICANCE = .0000
GAMMA = .39845
SOMERS'S D (ASYMMETRIC) = .14057 WITH FLARER DEPENDENT. = .13579 WITH AFS DEPENDENT.
SOMERS'S D (SYMMETRIC) = .14109

NUMBER OF MISSING OBSERVATIONS = 68

C.4 Historical Variables

FLAREHIS

FLARER	COUNT / ROW PCT / COL PCT / TOT PCT	NO FLARE OR FRST 0	C CLASS FLARE 1	M CLASS FLARE 2	X CLASS FLARE 3	ROW TOTAL
0 NO FLARE NXT DAY	COUNT	2336	972	382	28	3718
	ROW PCT	62.8	26.1	10.3	.8	82.9
	COL PCT	92.2	77.6	61.8	33.3	
	TOT PCT	52.1	21.7	8.5	.6	
1 C CLASS NEXT DAY	COUNT	176	228	160	22	586
	ROW PCT	30.0	38.9	27.3	3.8	13.1
	COL PCT	6.9	18.2	25.9	26.2	
	TOT PCT	3.9	5.1	3.6	.5	
2 M CLASS NEXT DAY	COUNT	21	48	67	25	161
	ROW PCT	13.0	29.8	41.6	15.5	3.6
	COL PCT	.8	3.8	10.8	29.8	
	TOT PCT	.5	1.1	1.5	.6	
3 X CLASS NEXT DAY	COUNT	0	4	9	9	22
	ROW PCT	0	18.2	40.9	40.9	.5
	COL PCT	0	.3	1.5	10.7	
	TOT PCT	0	.1	.2	.2	
COLUMN TOTAL		2533	1252	618	84	4487
		56.5	27.9	13.8	1.9	100.0

RAW CHI SQUARE = 784.63975 WITH 9 DEGREES OF FREEDOM. SIGNIFICANCE = 0
KENDALL#S TAU C = .17065. SIGNIFICANCE = 0
GAMMA = .60676
SOMERS#S D (ASYMMETRIC) = .21910 WITH FLARER DEPENDENT. = .43381 WITH FLAREHIS DEPENDENT.
SOMERS#S D (SYMMETRIC) = .29115

FIRSTAPP

FLARER	COUNT / ROW PCT / COL PCT / TOT PCT	ON DISK 0	FIRST TRANSIT 1	SECOND TRANSIT 2	THIRD TRANSIT 3	FOURTH TRANSIT 4	FIFTH TRANSIT 5	SIXTH TRANSIT 6	SEVENTH TRANSIT 7	ROW TOTAL
0 NO FLARE NXT DAY	COUNT	1391	1312	678	293	19	16	7	12	3718
	ROW PCT	37.4	35.3	18.2	7.6	.5	.4	.2	.3	82.9
	COL PCT	91.1	78.9	76.9	81.1	70.4	100.0	70.0	92.3	
	TOT PCT	31.0	29.2	15.1	6.3	.4	.4	.2	.3	
1 C CLASS NEXT DAY	COUNT	122	264	144	45	7	0	3	1	586
	ROW PCT	20.8	45.1	24.6	7.7	1.2	0	.5	.2	13.1
	COL PCT	8.0	15.9	16.3	12.9	25.9	0	30.0	7.7	
	TOT PCT	2.7	5.9	3.2	1.0	.2	0	.1	.0	
2 M CLASS NEXT DAY	COUNT	10	80	51	19	1	0	0	0	161
	ROW PCT	6.2	49.7	31.7	11.8	.6	0	0	0	3.6
	COL PCT	.7	4.8	5.8	5.4	3.7	0	0	0	
	TOT PCT	.2	1.8	1.1	.4	.0	0	0	0	
3 X CLASS NEXT DAY	COUNT	4	7	9	2	0	0	0	0	22
	ROW PCT	18.2	31.8	40.9	9.1	0	0	0	0	.5
	COL PCT	.3	.4	1.0	.6	0	0	0	0	
	TOT PCT	.1	.2	.2	.0	0	0	0	0	
COLUMN TOTAL		1527	1663	882	349	27	16	10	13	4487
		34.0	37.1	19.7	7.8	.6	.4	.2	.3	100.0

RAW CHI SQUARE = 163.71738 WITH 21 DEGREES OF FREEDOM. SIGNIFICANCE = .0000
KENDALL#S TAU C = .07674. SIGNIFICANCE = .0000
GAMMA = .27232
SOMERS#S D (ASYMMETRIC) = .08198 WITH FLARER DEPENDENT. = .10509 WITH FIRSTAPP DEPENDENT.
SOMERS#S D (SYMMETRIC) = .11565

PROTHIS

FLARER	COUNT / ROW PCT / COL PCT / TOT PCT	IND PARTI ICAL EVNT 0	PROTON 10 EVENT 1	GROUND EVENT 2	NO DATA 9	ROW TOTAL
NO FLARE NXT DAY	0	3695	23	3	0M	3718
		99.4	.5	.1	0	82.9
		83.5	35.1	60.0	0	
		82.4	.4	.1	0	
C CLASS NEXT DAY	1	578	15	1	0M	586
		97.3	2.6	.2	0	13.1
		12.9	26.3	20.0	0	
		12.7	.3	.0	0	
M CLASS NEXT DAY	2	140	19	1	1M	160
		87.5	11.9	.6	0	3.6
		3.2	33.3	20.0	0	
		3.1	.4	.0	0	
X CLASS NEXT DAY	3	19	3	0	0M	22
		86.4	13.6	0	0	.5
		.4	5.3	0	0	
		.4	.1	0	0	
	COLUMN TOTAL	4424	57	5	1M	4486
		98.6	1.3	.1	0	100.0

RAW CHI SQUARE = 198.43895 WITH 6 DEGREES OF FREEDOM. SIGNIFICANCE = 0
KENDALL#S TAU C = .02055. SIGNIFICANCE = .0000
GAMMA = .77662
SOMERS#S D (ASYMMETRIC) = .50210 WITH FLARER DEPENDENT. = .04649 WITH PROTHIS DEPENDENT.
SOMERS#S D (SYMMETRIC) = .08518

NUMBER OF MISSING OBSERVATIONS = 1

C.5 Total Sun Variables

FLUX

FLARER	COUNT / ROW PCT / COL PCT / TOT PCT	10 - 80 1	81-100 2	101-120 3	121-140 4	141-160 5	161-180 6	181-200 7	201-220 8	OVER 220 9	ROW TOTAL
NO FLARE NXT DAY	0	161	531	434	624	748	632	325	224	39	3718
		4.3	14.3	11.7	16.8	20.1	17.0	8.7	6.0	1.0	82.9
		92.5	87.6	83.0	82.9	81.5	79.2	83.8	81.8	75.0	
		3.6	11.8	9.7	13.9	16.7	14.1	7.2	5.0	.9	
C CLASS NEXT DAY	1	13	63	63	112	120	129	47	33	9	586
		1.7	10.8	10.8	19.1	20.5	22.0	8.0	5.6	1.5	13.1
		5.7	10.4	12.0	14.9	13.1	16.2	12.1	12.0	17.3	
		.2	1.4	1.4	2.5	2.7	2.9	1.0	.7	.2	
M CLASS NEXT DAY	2	3	9	24	17	44	34	13	14	3	161
		1.9	5.6	14.9	10.6	27.3	21.1	8.1	8.7	1.9	3.6
		1.7	1.5	4.6	2.3	4.8	4.3	3.4	5.1	5.8	
		.1	.2	.5	.4	1.0	.8	.3	.3	.1	
X CLASS NEXT DAY	3	0	3	2	1	6	3	3	3	1	22
		0	13.6	9.1	4.5	27.3	13.6	13.6	13.6	4.5	.5
		0	.5	.4	.1	.7	.4	.8	1.1	1.9	
		0	.1	.0	.0	.1	.1	.1	.1	.0	
	COLUMN TOTAL	174	606	523	754	918	798	388	274	52	4487
		3.9	13.5	11.7	16.8	20.5	17.8	8.6	6.1	1.2	100.0

RAW CHI SQUARE = 55.47394 WITH 24 DEGREES OF FREEDOM. SIGNIFICANCE = .0003
KENDALL#S TAU C = .03660. SIGNIFICANCE = .0000
GAMMA = .10937
SOMERS#S D (ASYMMETRIC) = .03216 WITH FLARER DEPENDENT. = .09305 WITH FLUX DEPENDENT.
SOMERS#S D (SYMMETRIC) = .04780

C.6 Events During This 24 Hours

FLARERT

FLARER	COUNT / ROW PCT / COL PCT / TOT PCT	NO FLARE TODAY 0	C CLASS TODAY 1	M CLASS TODAY 2	X CLASS TODAY 3	NO DATA 9	ROW TOTAL
0 NO FLARE NXT DAY	COUNT	3311	337	66	3	1M	3717
	ROW PCT	89.1	9.1	1.8	.1	0	82.9
	COL PCT	90.3	54.0	38.2	13.6	0	
	TOT PCT	73.8	7.5	1.5	.1	0	
1 C CLASS NEXT DAY	COUNT	307	219	55	5	0M	586
	ROW PCT	52.4	37.4	9.4	.9	0	13.1
	COL PCT	8.4	35.1	31.8	22.7	0	
	TOT PCT	6.8	4.9	1.2	.1	0	
2 M CLASS NEXT DAY	COUNT	47	60	47	7	0M	161
	ROW PCT	29.2	37.3	29.2	4.3	0	3.6
	COL PCT	1.3	9.6	27.2	31.8	0	
	TOT PCT	1.0	1.3	1.0	.2	0	
3 X CLASS NEXT DAY	COUNT	2	8	5	7	0M	22
	ROW PCT	9.1	36.4	22.7	31.8	0	.5
	COL PCT	.1	1.3	2.9	31.8	0	
	TOT PCT	.0	.2	.1	.2	0	
COLUMN TOTAL		3667	624	173	22	1M	4486
		81.7	13.9	3.9	.5	0	100.0

RAW CHI SQUARE = 1400.44238 WITH 9 DEGREES OF FREEDOM. SIGNIFICANCE = 0
KENDALL#S TAU C = .17043. SIGNIFICANCE = 0
GAMMA = .77812
SOMERS#S D (ASYMMETRIC) = .41107 WITH FLARER DEPENDENT. = .43317 WITH FLARERT DEPENDENT.
SOMERS#S D (SYMMETRIC) = .42183

NUMBER OF MISSING OBSERVATIONS = 1

RECSPOT

FLARER	COUNT / ROW PCT / COL PCT / TOT PCT	LT 10 PERCENT 1	10 TO 20 PERCENT 2	20 TO 30 PERCENT 3	30 TO 50 PERCENT 4	50 TO 65 PERCENT 5	66 TO100 PERCENT 6	OVER 100 PERCENT 7	OVER 200 PERCENT 8	NEVER OCCURRED 98	NO DATA 99	ROW TOTAL
0 NO FLARE NXT DAY	COUNT	2817	517	160	72	62	19	40	16	0M	7M	3703
	ROW PCT	76.1	14.0	4.3	1.9	1.7	.5	1.1	.4	0	0	82.9
	COL PCT	92.1	77.6	60.6	57.1	47.0	47.5	38.8	20.8	0	0	
	TOT PCT	63.0	11.6	3.6	1.6	1.4	.4	.9	.4	0	0	
1 C CLASS NEXT DAY	COUNT	212	119	84	39	53	13	37	26	1M	2M	583
	ROW PCT	36.4	20.4	14.4	6.7	9.1	2.2	6.3	4.5	0	0	13.0
	COL PCT	6.9	17.9	31.8	31.0	40.2	32.5	35.9	33.8	0	0	
	TOT PCT	4.7	2.7	1.9	.9	1.2	.3	.8	.6	0	0	
2 M CLASS NEXT DAY	COUNT	30	27	19	12	15	4	22	31	0M	1M	160
	ROW PCT	18.8	16.9	11.9	7.5	9.4	2.5	13.7	19.4	0	0	3.6
	COL PCT	1.0	4.1	7.2	9.5	11.4	10.0	21.4	40.3	0	0	
	TOT PCT	.7	.6	.4	.3	.3	.1	.5	.7	0	0	
3 X CLASS NEXT DAY	COUNT	1	3	1	3	2	4	4	4	0M	0M	22
	ROW PCT	4.5	13.6	4.5	13.6	9.1	18.2	18.2	18.2	0	0	.5
	COL PCT	.0	.5	.4	2.4	1.5	10.0	3.9	5.2	0	0	
	TOT PCT	.0	.1	.0	.1	.0	.1	.1	.1	0	0	
COLUMN TOTAL		3060	666	264	126	132	40	103	77	9M	10M	4468
		68.5	14.9	5.9	2.8	3.0	.9	2.3	1.7	0	0	100.0

RAW CHI SQUARE = 1144.88504 WITH 21 DEGREES OF FREEDOM. SIGNIFICANCE = 0
KENDALL#S TAU C = .19592. SIGNIFICANCE = 0
GAMMA = .68651
SOMERS#S D (ASYMMETRIC) = .29232 WITH FLARER DEPENDENT. = .49845 WITH RECSPOT DEPENDENT.
SOMERS#S D (SYMMETRIC) = .36852

NUMBER OF MISSING OBSERVATIONS = 19

DATE
ILMED
-8